

FOR OFFICIAL USE ONLY

ACCESS DB #

25108

PLEASE PRINT CLEARLY

Location (Bldg/Room#):

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: Requester's Full Name: Examiner #:

Art Unit: Phone (30\_) Serial Number:

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention:

Inventors (please provide full names):

Earliest Priority Date:

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

\*For Sequence Searches Only\* Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: HART

Searcher Phone #: 305-9203

Searcher Location: 12214

Date Searcher Picked Up: 2/9/00

Date Completed:

Searcher Prep & Review Time:

Online Time:

Type of Search

2 NA Sequence (#)

3 AA Sequence (#)

Structure (#)

Bibliographic

Litigation

Fulltext

Other

Vendors and Cost

STN Dialog

Questel/Orbit Dr. Link

Lexis/Nexis Westlaw

WWW/Internet

X In-house sequence systems (list)

Other (specify)

ADSS 82

25108  
**STIC-Biotech/ChemLib**

---

**From:** Duffy, Patricia  
**Sent:** Tuesday, February 08, 2000 1:51 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence Search 09/060,609

In RE: 09/060,609

Please search SEQ ID NO:1 and a oligomer search of SEQ ID NO:1.  
Please search SEQ ID NO:2 and an oligomer search of SEQ ID NO:2.  
Please backtranslate the amino acid sequence of SEQ ID NO:2 into NA and run NA.

Please include an interference search.

Thanks,

**Patricia A. Duffy**

**Art Unit 1645; CM-1 8D05**

**305-7555**



```
alignment_scores:
  Quality: 306.50      Length: 184
  Ratio: 2.620         Gaps: 7
  Percent Similarity: 63.587  Percent Identity: 38.043

alignment_block:
  US-09-060-609-2 x AC007175/rev

Align seg 1/1 to reverse of: AC007175 from: 1 to: 188633

107 SerLeuLysCysGluAspLeuLys...ValGlyGlnTyrIleCysLysAsp 122
127457 ATATGGAGTCAACAGAGTCCAGATGATGGCCAGTTCATGTGCCCCGA 127408
122 pPro.....LysIleAsnAspAlaThrGlnGluProValAspCyst 136
27407 CCCGGGAGGGTCAATCATGATCCAGACACAGCAGTGGCGGATGCA 127358
136 hrAsnTyr...ThrAlaHisValSerCysPheProAlaProAsnIleThr 151
127357 CCAGGAGGGTCCGCGCGCTGTGTGATCGCCGCCAACGAGATCAAC 127308
152 CysLysAspSerSerGlyAsnGlnThrHisPheThrGlyAsnGluValGI 168
127307 TGCACGGAG...ACTGGCAATGCCAG..... 127284
168 yPhePheLysProIleSerCysArg...AsnVal..... 178
127283 .TTCACCCGAGAGTGCCTCGAAGTGGAGCGTGACGTGCAGTACATATC 127235
179 ..... 179
127234 TCATAAACATAATCTTAAGACCTATATCTTATCTTATAGGAATG 127185
180 lYtyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeu 196
127184 GCTATCACCTGGACACACCACTGCTACTCTGTCTTTCTGGCGATGTC 127135
197 GlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysph 213
127134 GCGGTGGATAGATCTTATTTGGCTATTCGGGATCGGACTCCTCAAGTT 127085
213 eCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuI 230
27084 CTGCACCTCGCGGATCTTCTGGCCAGCTGATGACATCGTCTGA 127035
230 leSerMetGlnIleValGlyProSerAspGlySerTyrIleIleAsp 246
127034 TAGCCCTGCAGGTGTGGTTCGGCGGATGGCTCCGCTATGTGATACC 126985
247 TyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPhe 263
126984 TACTACGGAGCGGCATCCATCCTGTCGCGCAGCAGCAACACACGATCCG 126935
263 g 263
126934 G 126934

seq_name: em_htgl:AC004434
seq_documentation_block:
ID AC004434 standard; DNA; HTG: 83204 BP.
XX AC AC004434; AC003531; AC003532; AC003533; AC003838;
XX AC AC004434.4
SV AC004434.4
XX
DT 18-MAR-1998 (Rel. 55, Created)
DT 25-MAR-1999 (Rel. 59, Last updated, Version 17)
XX
DE Drosophila melanogaster; Chromosome 2R; Region 57D11-57E2; p1 clone
```

```
DE DS01261, WORKING DRAFT SEQUENCE, 10 unordered pieces.
XX HTG: HTGS_PHASE1.
XX Drosophila melanogaster (fruit fly)
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC Drosophilidae; Drosophila.
XX [1]
RN 1-83204
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
  Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
  Farfan D.E., Galle R., George R.A., Hummasti S.R., Kim E., Hoskins R.A.,
  Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim S.H., Lee B.,
  Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
  Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
  Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
  Zieran L.L., Rubin G.M.;
  "Sequencing of Drosophila melanogaster";
  Unpublished.
RN [2]
RP 1-83204
RA Celniker S.E., George R.A., Galle R., Svirskas R.R., Hoskins R.A.,
  Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Chavez C., Chew M.,
  Doyle C.M., Farfan D.E., Flanagan J., Houston K.A., Hummasti S.R.,
  Karra K., Kearney L., Kim S.H., Lee B., Lomotan M.A., Mak J., Mazda P.,
  Mok M.S., Moshrefi A.R., Moshrefi M., Nixon K., Pacleb J.M., Park S.,
  Pfeiffer B., Poon L., Sequeira A., Svirskas R.R., Wan K.H.,
  Zhang R., Zieran L.L., Kimmel B.E.;
  Submitted (17-MAR-1998) to the EMBL/GenBank/DBJ databases.
RL Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121,
  Berkeley, CA 94720, USA
XX

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 400 bases, phrap computed error
rate <= 1/10. Pl library location: 14-13.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 738: contig of 738 bp in length
* 739 818: gap of unknown length
* 819 1918: contig of 1100 bp in length
* 1919 1998: gap of unknown length
* 1999 2644: contig of 646 bp in length
* 2645 2724: gap of unknown length
* 2725 3214: contig of 490 bp in length
* 3215 3294: gap of unknown length
* 3295 4442: contig of 1148 bp in length
* 4443 4522: gap of unknown length
* 4523 5169: contig of 647 bp in length
* 5170 5249: gap of unknown length
* 5250 6692: contig of 1443 bp in length
* 6693 6772: gap of unknown length
* 6773 9720: contig of 2948 bp in length
* 9721 9800: gap of unknown length
* 9801 13304: contig of 3504 bp in length
* 13305 13385: gap of unknown length
* 13386 83204: contig of 69820 bp in length.
Key Location/Qualifiers
source 1. .83204
```







```

* 137260 137339: gap of unknown length
* 137340 137991: contig of 652 bp in length
* 137992 138071: gap of unknown length
* 138072 138762: contig of 691 bp in length
* 138763 138842: gap of unknown length
* 138843 139393: contig of 551 bp in length
* 139394 139473: gap of unknown length
* 139474 140102: contig of 629 bp in length
* 140103 140182: gap of unknown length
* 140183 140811: contig of 629 bp in length
* 140812 140891: gap of unknown length
* 140892 141500: contig of 609 bp in length
* 141501 141580: gap of unknown length
* 141581 142262: contig of 682 bp in length
* 142263 142342: gap of unknown length
* 142343 143046: contig of 704 bp in length
* 143047 143126: gap of unknown length
* 143127 143812: contig of 686 bp in length
* 143813 143892: gap of unknown length
* 143893 144614: contig of 722 bp in length
* 144615 144694: gap of unknown length
* 144695 145373: contig of 679 bp in length
* 145374 145453: gap of unknown length
* 145454 146116: contig of 663 bp in length
* 146117 146196: gap of unknown length
* 146197 146808: contig of 612 bp in length
* 146809 146888: gap of unknown length
* 146889 147514: contig of 626 bp in length
* 147515 147594: gap of unknown length
* 147595 148161: contig of 567 bp in length
* 148162 148241: gap of unknown length
* 148242 148927: contig of 686 bp in length
* 148928 149007: gap of unknown length
* 149008 149592: contig of 585 bp in length.

FEATURES
      Location/Qualifiers
      1..149592
            /organism="Drosophila melanogaster"
            /strain="y; cn bw sp"
            /db_xref="taxon:7227"
            /clone="DS02336 (D440)"
            /chromosome="2"
            /clone_lib="pl library, partial Sau3A in pNS
            /map="60C8-60D2"
BASE COUNT 37500 a 33853 c 33980 g 38891 t 5368 others
ORIGIN

alignment_scores:
      Quality: 257.00      Length: 85
      Ratio: 3.725        Gaps: 0
      Percent Similarity: 81.176      Percent Identity: 54.118

alignment_block:
US-09-060-609-2 x AC005718/rev

Align seg 1/1 to reverse of: AC005718 from: 1 to: 149592

179 AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195
||||| : : : : :
25592 AATGGCTATACCTGGACACACACTGCTACTCTCTGCTTTCTGGGCAT 25543

195 pleuGlyAlaaspargPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuL 212
: : : : :
25542 GTTCGGCTGGATAGATTCTATTTTGGGTATCCGGGGCATCCGACTCCCTCA 25493

212 ysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228
||||| : : : : :
25492 AGTTCTGCACCTCGCGCGGATGTTCTTGGGCCAGCTGATTGCATCGTG 25443

229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIl 245
||||| : : : : :
25442 CTGATAGCCCTGCAGGTTGTGGGTCGGCGGATGCTCGGCTATGTGAT 25393

```

```

245 eASPTyrrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrP 262
| |||||::: :::: ||| |||
25392 ACCCTACTACGAGCGGCATCCATCCATCGTCGCGGACACACACCGT 25343
262 heArg 263
::|||
25342 ACCGG 25338
seq_name: gb_htgl:DMBR30C13
seq_documentation_block:
LOCUS DMBR30C13 117674 bp DNA HTG 19-OCT-1999
DEFINITION Drosophila melanogaster chromosome X clone BACR30C13 map 8D strain
Y: cn bw sp, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AL122025
VERSION AL122025.1 GI:6090476
KEYWORDS HTG: HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 117674)
AUTHORS Madueno,E., de Pablos,B. and Modolell,J.
TITLE Sequencing the X chromosome of Drosophila melanogaster
JOURNAL Unpublished
REMARK Centro de Biologia Molecular Severo Ochoa, Consejo Superior de
Investigaciones Cientificas and Universidad Autonoma de Madrid,
28049 Madrid, Spain.
2 (bases 1 to 117674)
AUTHORS Benos,P.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) European Drosophila Genome Sequencing
Consortium
COMMENT This is a 'working draft' sequence. It currently
consists of 153 contigs. The true order of the pieces is not known
and their order in this sequence record is arbitrary. Gaps between
the contigs are represented as runs of N, but the exact sizes of
the gaps are unknown. This record will be updated with the finished
sequence. 1 513: contig of 513 in length
514 613: gap of unknown length
614 1380: contig of 767 in length
1381 1480: gap of unknown length
1481 1995: contig of 515 in length
1996 2095: gap of unknown length
2096 2862: contig of 767 in length
2863 2962: gap of unknown length
2963 3734: contig of 792 in length
3735 3854: gap of unknown length
3855 4657: contig of 803 in length
4658 5330: contig of 573 in length
5331 5430: gap of unknown length
5431 5932: contig of 502 in length
5933 6032: gap of unknown length
6033 6474: contig of 442 in length
6475 6574: gap of unknown length
6575 7188: contig of 614 in length
7189 7288: gap of unknown length
7289 7513: contig of 225 in length
7514 7613: gap of unknown length
7614 8268: contig of 655 in length
8269 8368: gap of unknown length
8369 9346: contig of 978 in length
9347 9447 10066: contig of 620 in length
10067 10166: gap of unknown length
10167 10814: contig of 648 in length
10815 10915: gap of unknown length
10915 11203: contig of 289 in length
11204 11303: gap of unknown length
11304 11941: contig of 638 in length
11942 12041: gap of unknown length
12041: gap of unknown length
12042 12655: contig of 624 in length
12656 12765: gap of unknown length
12766 13410: contig of 645 in length
13411 13510: gap of unknown length
13511 13996: contig of 486 in length
13997 14096: gap of unknown length
14097 14677: contig of 581 in length
14678 14777: gap of unknown length
14778 15318: contig of 541 in length
15319 15418: gap of unknown length
15419 16047: contig of 629 in length
16048 16147: gap of unknown length
16148 16788: contig of 641 in length
16789 16888: gap of unknown length
16889 17502: contig of 614 in length
17503 17602: gap of unknown length
17603 18281: contig of 679 in length
18282 18381: gap of unknown length
18382 18819: contig of 438 in length
18819 18919: gap of unknown length
18920 19754: contig of 835 in length
19755 19854: gap of unknown length
19855 20699: contig of 845 in length
20700 20799: gap of unknown length
20800 21833: contig of 384 in length
21834 21883: gap of unknown length
21884 22021: contig of 738 in length
22022 22121: gap of unknown length
22122 23187: contig of 1066 in length
23188 23287: gap of unknown length
23288 23946: gap of unknown length
23946 24046: gap of unknown length
24046 24616: contig of 571 in length
24617 24716: gap of unknown length
24717 25676: contig of 960 in length
25677 25776: gap of unknown length
25777 26190: contig of 414 in length
26191 26290: gap of unknown length
26291 27392: contig of 1002 in length
27393 27392: gap of unknown length
27393 28107: contig of 715 in length
28108 28207: gap of unknown length
28208 28729: contig of 522 in length
28729 28829: gap of unknown length
28830 29976: contig of 1147 in length
29977 30076: gap of unknown length
30077 30928: contig of 852 in length
30929 31028: gap of unknown length
31029 31903: contig of 875 in length
31904 32003: gap of unknown length
32004 32418: contig of 415 in length
32419 32518: gap of unknown length
32519 33132: contig of 614 in length
33133 33232: gap of unknown length
33233 33923: contig of 591 in length
33924 34827: contig of 904 in length
34828 34927: gap of unknown length
34928 35934: contig of 1007 in length
35935 36034: gap of unknown length
36035 36962: contig of 928 in length
36963 37062: gap of unknown length
37063 37679: contig of 617 in length
37680 37779: gap of unknown length
37780 38484: contig of 705 in length
38485 38584: gap of unknown length
38585 39541: contig of 957 in length
39542 39641: gap of unknown length
39642 40517: contig of 876 in length
40518 40617: gap of unknown length
40618 41307: contig of 690 in length
41308 41407: gap of unknown length
41408 42011: contig of 604 in length

```

42012	42111: gap of unknown length	
42112	42828: contig of 717 in length	
42829	42928: gap of unknown length	
42929	43593: contig of 665 in length	
43594	43693: gap of unknown length	
43694	44391: contig of 698 in length	
44391	44491: gap of unknown length	
44491	45147: contig of 856 in length	
45147	45247: gap of unknown length	
45248	46798: contig of 1551 in length	
46799	46898: gap of unknown length	
46899	47830: contig of 932 in length	
47831	47931: gap of unknown length	
47931	48883: contig of 953 in length	
48884	48983: gap of unknown length	
48984	50341: contig of 1358 in length	
50342	50441: gap of unknown length	
50442	51797: contig of 1356 in length	
51798	51897: gap of unknown length	
51898	52237: contig of 340 in length	
52238	52337: gap of unknown length	
52338	53710: contig of 1373 in length	
53711	53810: gap of unknown length	
53811	54430: contig of 620 in length	
54431	54530: gap of unknown length	
54531	55595: contig of 1065 in length	
55596	55695: gap of unknown length	
55696	56555: contig of 860 in length	
56556	56655: gap of unknown length	
56656	57909: contig of 1254 in length	
57910	58009: gap of unknown length	
58010	59336: contig of 1427 in length	
59337	59536: gap of unknown length	
59537	60724: contig of 1188 in length	
60725	60824: gap of unknown length	
60825	62193: contig of 1369 in length	
62194	62293: gap of unknown length	
62294	63069: contig of 776 in length	
63070	63169: gap of unknown length	
63170	64233: contig of 1064 in length	
64234	64333: gap of unknown length	
64334	64782: contig of 449 in length	
64783	64882: gap of unknown length	
64883	66244: contig of 1362 in length	
66245	66344: gap of unknown length	
66345	67507: contig of 1163 in length	
67508	67607: gap of unknown length	
67608	68877: contig of 1270 in length	
68878	68977: gap of unknown length	
68978	70692: contig of 1715 in length	
70693	70792: gap of unknown length	
70793	71673: contig of 981 in length	
71674	71773: gap of unknown length	
71774	72809: contig of 1036 in length	
72810	72909: gap of unknown length	
72910	74298: contig of 1389 in length	
74299	74398: gap of unknown length	
74399	76722: contig of 2324 in length	
76723	76822: gap of unknown length	
76823	79379: contig of 2557 in length	
79380	79479: gap of unknown length	
79480	79862: contig of 383 in length	
79863	79962: gap of unknown length	
79963	80475: contig of 513 in length	
80476	80575: gap of unknown length	

alignment\_scores:  
 Quality: 183.50 Length: 188  
 Ratio: 1.799 Gaps: 9  
 Percent Similarity: 54.255 Percent Identity: 32.979  
 alignment\_block:  
 US-09-060-609-2 x DMBR30C13

Align seg 1/1 to: DMBR30C13 from: 1 to: 117674

61 SerLysMetalAlaAlaThrProSerGlyProSerAlaProGluAlaVa 77  
 4022 TCAACATGTCAGTGGCGAAGCGAGTTCACGACAAATGAAT 4071  
 77 lThrAlaArgLeuValGlyValLeuTrrPheValSerValThrGly. 93  
 4072 T.....GCCGATACGCTA 4085  
 94 ProTrpGlyAlaValAlaThr.SerAlaGlyGlyGluSerLeuLysC 110  
 4086 CCAACGGAATGTGGCAACAGCTGTGGCAGCGTCCAGT..... 4128  
 110 ysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsn 126  
 4129 .....TGCAAC 4134  
 127 AspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHis.... 141  
 4135 TCGCGCAGGATAAGCTATTCCGT...ACCAACTGCACCGTGACCATGA 4181  
 142 .ValSerCysPheProAlaProAsnIleThrCysLysAspSerSerGlyA 158  
 4182 TGTGCTCTGCCTC.....GGCA 4198  
 158 snGluThrHisPheThrGlyAsnGluValGlyPhePheLysProIleSer 174  
 4199 ATCGATCC...TTTACCGCGAAC.....CTGGCA 4224  
 175 CysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLe 191  
 4225 TGCAACTGGACGAGGCTATCGGTGGAGCACAGCCCTGCTGATTAGCCT 4274  
 191 uPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaL 208  
 4275 CACGCTAGGGGCTTGGAGCGCATCGATTCTACTTGGGCCACTGGCAGG 4324  
 208 euGlyLeuLysPheCysThrValGlyPheCysGlyIleGlySerLeu 224  
 4325 AAGGATGGAAAGCTTCAGCTTTGGCGCCTCGGCGTCTGGACCATC 4374  
 225 IleAspPheIleLeuIleSerMetGlnIleValGlyProSerAspGlyse 241  
 4375 ATCGATGCTGCTCACTCGATGCAITATTGGGGCCAGCGGATGGCTC 4424  
 241 rSerTyrIle 244  
 4425 ACTTTACATA 4434  
 seq\_name: gb\_in2:AF181623  
 seq\_documentation\_block:  
 LOCUS AF181623 1440 bp mRNA INV 16-SEP-1999  
 DEFINITION Drosophila melanogaster BCDNA.GH02340 (BCDNA.GH02340) mRNA,  
 complete cds.  
 ACCESSION AF181623  
 VERSION AF181623.1 GI:5901801  
 KEYWORDS FLI.CDNA  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1440)  
 AUTHORS Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,  
 Tsang,G., Agbayani,A., Arcana,F.T., Baxter,E., Blazej,R.G.,  
 Butenhoof,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,  
 Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,  
 Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,  
 Snir,E., Svirska,R.R., Weinburg,T. and Celnikier,S.E.



NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 5' cosmid is c11d9, 3100 bp overlap. Actual start of this cosmid is at base position 1 of CELC41d11; actual end is at 40912 of CELC41d11

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

#### FEATURES

##### source

Location/Qualifiers  
1. 40912  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="I"  
/clone="C41d11"  
complement(6009. .8933)  
/gene="C41d11.3"  
complement(join(6009. 6173,6597. 6821,7005. .7699, 7830. 8157,8748. .8933))  
/gene="C41d11.3"

/note="weak similarity to rat mitochondrial capsule selenoprotein (GB:X87883) and integrin beta oligodendroglia chain precursor (NID:g520566); coded for by C. elegans cDNA yk2d8.3; coded for by C. elegans cDNA yk130f.3; coded for by C. elegans cDNA yk101b2.3; coded for by C. elegans cDNA yk121el.3; coded for by C. elegans cDNA yk130f.5; coded for by C. elegans cDNA yk101b2.5; coded for by C. elegans cDNA yk121el.5"

/codon\_start=1

/protein\_id="AAC48138.1"

/db\_xref="GI:2103493"

/translation="MVPLVKKQMSIDEKIMEDDEDEFGDSMTIKRGGSSLRSLRSLKSPKTDQKPGADTEKSCSTPLTAPLLSRQGSTSLRFRVSPSTFLKPIATPIIDVSPTRPDPPELRIIVAAAPVSTELSEPTSKTPEKAPAPAMOKSALKRHHAFNSVREIFIPGIRKIRFESGLHVHVDREQGSTVPTGEDISLDMHKNHSHRYFSLSGKRPIILLIYADDLSGVEDIDPSDEDRIDYFASAKSIPRINORHRIKLLKSGVAKENTDAISLRNTRVEGSCENGVCPLPETCOAIDGHCQVDGGEWPTQPCA CFAETCNPEGRTPDEAVHEFRQRTIMNWRASQKTSIGSVYTKRFAFDSDEEDKA KTWLLKNAIPKLEESTPKVPTVYTRRSRSLSDIRESTSETSGTSTPSTLSERI KAMHLDNVEQVEEKEDDADVFMDNKEDTIIIEADLTLOEYVVEQDMVAVTESSE TTIDANEMIEATAMSPVKTLLV"

complement(18663. .20974)

/gene="C41d11.4"

complement(join(18663. 18779,18829. 18944,19622. .19868, 20561. .20974))

/gene="C41d11.4"

/note="coded for by C. elegans cDNA yk21el.5; coded for by C. elegans cDNA yk2d8.5"

/codon\_start=1

/protein\_id="AAC48139.1"

/db\_xref="GI:2103494"

/translation="MVSVAASLKRAGVPSLRTTTSALPIKMKKPMVPLRIEDQHOTPIRKLAESPISLEDQTPPMRRSSGMLLRQPERIRROEEDVETPGSPVKRRRS GSMPLRQAQSDVPIRMEIGNSLFPVLRKRRTSLTCLFSDMPPIITPQRLRGIS ETATIDDDVFVKSPKTPITSPKRVQDSNRSTSPKRRYISSFSIRLEBEKEI EKAVPTPTMSTPMRKFSTIFNPRAVTERIAQTYGTGSRQLKSTSTSVLPAPPALVA RSTSGEYIYSATEGLAFFKY"

#### gene

#### CDS

21167. .23273  
/gene="C41d11.2"  
join(21167. .21266,21736. .21863,21909. .22031,22078. .22273, 22321. .22420,22470. .22941,23007. .23273)  
/gene="C41d11.2"  
/note="coded for by C. elegans cDNA yk83cl.3; coded for by C. elegans cDNA yk83cl.5"  
/codon\_start=1  
/protein\_id="AAC48137.1"  
/db\_xref="GI:2105492"  
/translation="MVSEINLQLEFNMHLVILNLQMSKASLFKNWLSKMTAVTITAPSVKHLISLVVKKLVKVVHSELHAGISEVSGDAGCVLTGLVLEDSRLIETNC FPTVNEPEVDDDAANAQOYEQKHQHEMLDMLRLETMINDIYEIYQSHQFGAGFS HDLVESMEFYQAMPENYVLIYDKTRQGLSLRAWRLSTALDASNDWRPELVK AAGLYQNMFEELPIIKSSYNNLVMSLAKSSCDKYSIRHFDLGSKSLKSV RAMANDVLAKSIQSLATYDKORHDNMWFSLTOKRVIRNFETKILNIVLSLKFIF RSVTAFKPKVLFFENHFTYRGSTNYKTCGEGYIFGGIKQENESVARGDPTIPDD IKRKAPQQLTRNGLLDELLASFDNALADSKVTISENITKMFIAEVAEAEKRVAGTK DRTLSVSSSTR"  
complement(23679. .27940)  
/gene="C41d11.5"  
complement(join(23679. .24030,24286. .24440,24497. .24619, 25129. .25305,25352. .25478,25755. .26011,26081. .26330, 26721. .26791,27033. .27433,27495. .27600,27647. .27731, 27783. .27940))  
/gene="C41d11.5"

#### gene

#### CDS

complement(join(23679. .24030,24286. .24440,24497. .24619, 25129. .25305,25352. .25478,25755. .26011,26081. .26330, 26721. .26791,27033. .27433,27495. .27600,27647. .27731, 27783. .27940))  
/gene="C41d11.5"  
/note="similar to the DNA/RNA non-specific endonuclease family; coded for by C. elegans cDNA CEESU20R"  
/codon\_start=1  
/protein\_id="AAC48141.1"  
/db\_xref="GI:2105496"

/translation="MMSLSLNNENLDEIIDFESKCRFPVMKRAVCITHSSLOFVD LFMONLMEHSGFSTPIIARQDETEVLKTMSEVKINERETIRSHRGNVSEVA IFIGTCPVRSYTSYRVPITICAEADSTHSCGDTCAELNDFERKINRLFMHSCSKE QSNSTNKNHRLVFVFKGSDVMNEIEBELVHEFPNFFELFFFRDPFGSKQ SFHYFIHEKNCNDFVAPNDGKWLVPVYIMLHKILFLICLASEIPTIGSISGTGD VKSKNCDGSAGLTTFPGDCRTGVKVNCTSRKGNPNVNNNVAFCVQWLLPG DYDCEPATNCSSTKLLTKCSAHSVITCMGORNFYKPCNWSGSGWTQMILSV VLGGAGDRFVLGLKSAIGLFSFGLGVTLDVLLVAVGYIKPDGASIRNMIGKV AGRATAIGISLAGKYSNDDDLIFRNVQSATNPMNQIQVSEPTMKVPSLADAMGP SRSAIMKHGYPGTNRYTDEVLSDYKTAHWVCEHLTPERLUKHAEGVDKLIKE FKPDITFFQKLSQNTDYKCSGDRGLAAAGNHRKSQLAVQTFHLSKSPQVGRGF NRKNDWIDEMHCRRAKKMINSYIITGPLYLPKLEGDKKYIKYQVIGNNNVAVPTFH FKVALFEVTPGKFELESYILPNAVEDTVEISKHFVPLDAVERSAGLEIFARLDPKSI VKENGAKKGGLN"

complement(28522. .33822)

/gene="C41d11.7"

complement(join(28522. .28843,29286. .29418,29463. .29705, 30317. .30489,30536. .30718,31647. .31861,31908. .32009, 32388. .32461,32961. .33352,33758. .33822))

/gene="C41d11.7"

/note="similar to mammalian DNA-binding proteins SMUBP-2"

/codon\_start=1

/evidence="not experimental"

/protein\_id="AAC48142.1"

/db\_xref="GI:2105497"

/translation="MIEADKFELELVKKGVLSVGLNQKQNTNPVQYLTEDNPESLT LQDVLESHMTDPIELNSDMEKEAGIPESNEKQRTAIRKALNENRKLVCIQPGPTG KFTVTLTLCLLQIQKQVYLAFTREALNTMMTKTLKRMKIVHEHADMFTNEY RDVINKSDRALMAAEVRLKAFNGEITENVLDMQSIINRVNEVGAIVGNVR VAFATIGASFVDFVNMKHKFDCICIIDEAQVMEATWPAVKMKRIVMAGDPKOLP ALVFTDEAKAFLQNSVMDRIELKKNFSWIMLENGYRKNATATNSCTFYHQLKT DYKHEYLSLHTILNPQPKFRNLDFDLVIDTSLDERKLETIHVAFTDTSINKT KQGSYANLAEAKIAGHYORLLKYGVQPSDAITPPYQGLSLTKVIMEEFGAETGY TDFVOTTICTVDVQGVKEVYVFTVMRNPRTMGFVSELRLNVIIVRAKRFMFYI GNGYLLAESKKDEIRKLYDCFKNAKRPHNNAFGDGNVTDVSNNGQDLESFMYI SNDEMINWCKREFENGPDPYAKKALAKEETRLMKTISKVKIYKRGSTQLASE LDQLDCSSN"

34408. .35688

/gene="C41d11.1"

join(34408. .34592,34661. .34754,34805. .34917,34967. .35153, 35210. .35348,35399. .35505,35551. .35688)

/gene="C41d11.1"

#### gene

#### CDS

```

/codon_start=1
/evidence=not_experimental
/protein_id="AAC48136.1"
/db_xref="GI:2105491"
/translation="MTDKENFYSGFLRPLSGDGRFLIKVACFPQKERNVVELT
TDTKREDYGEVHFVKGKPLQNGKSVFPEENYETFFHENKMFDRIVHTDQDFVVG
VNDPLSDIYIHDPDLCKMIDTRISIVTPGKKYNEITFEVKONNNSDIEIVGGEVQK
EKNSPSEAKETIQMEQVHEVMKVEYKNEETANNDVIGHLLKLIKFDKDLTSG
TVVLLGNDRHKFERLFKREDYFRLSNNTLTLHKFWQYGLEKI"
gene
complement(37163..39721)
/genes="C41D11.6"
CDS
complement(37163..39721)
/genes="C41D11.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC48140.1"

```

```

alignment_scores:
  Quality: 162.00      Length: 159
  Ratio: 2.051        Gaps: 5
  Percent Similarity: 49.686  Percent Identity: 28.931

alignment_block:
US-09-060-609-2 x CELC41D11/rev ..

Align seg 1/1 to reverse of: CELC41D11 from: 1 to: 40912
110 CysGluAspLeuLysValGlyClnTyrIleCysLysAspProLysIleAs 126
|||||.....|.....|.....|.....|.....|.....|.....|
26170 TGTGGCAACTCTACCTGGGATATGATTC...GAGCCAGCAGCAAA 26124

126 nAspAlaThrGlnGluPro.....ValAsnCysThr..... 136
|.....|.....|.....|.....|.....|.....|.....|
26123 TTGCTCCAGAGTAGTACAAAGCTCTCGTGACAAAGTGCTCGGTGACCA 26074

136 ..... 136

```

```

26073 ATTCTTTAATTAATATCTGTTTAAATAACCTAATTTCAAAAAA 26024
137 ...AsnTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCy 152
|||||.....|.....|.....|.....|.....|.....|.....|
26023 TTGAATTTTCAGGCTCACAGCTCA.....GTAATATG 25992

152 sLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyP 169
|.....|.....|.....|.....|.....|.....|.....|
25991 TATG.....GGACAACGTAAT 25975

169 hePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysVal 185
|||||.....|.....|.....|.....|.....|.....|.....|
25974 TCTACAAGCGAATTCATGTAACTGGTCATCTGGATACAGTTGGACAAA 25925

186 AlavalAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTy 202
|||||.....|.....|.....|.....|.....|.....|.....|
25924 ACAATGATCTCTCATGCTCGCGCGGTTTGGAGCTGATCGTTTTCTA 25875

202 rLeuGlyTyrProAlaLeuGlyLeuLysPheCysThrValGlyPheC 219
|||||.....|.....|.....|.....|.....|.....|.....|
25874 CTGGGACTCTGGAATCTGAAATCGAATGGAAACCTTTTCAGTTTCGGTGT 25825

219 ysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleVal 235
|||||.....|.....|.....|.....|.....|.....|.....|
25824 TTGGTGTGGAGCTCTCTGGACGCTGTCTCATCGCAGCTGGATGACATT 25775

236 GlyProSerAspGlySerSerTyrIle 244
|||.....|.....|.....|.....|.....|.....|.....|
25774 AAACCTTACGATGATGATGATATT 25748

```

seq\_name: gb\_htg6:AC016672

seq\_documentation\_block:

LOCUS AC016672 205157 bp DNA 04-DEC-1999  
 DEFINITION Homo sapiens clone RP11-14C10, WORKING DRAFT SEQUENCE, 1 unordered

```

ACCESSION AC016672
VERSION AC016672.1 GI:6524410
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 205157)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205157)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 205157: contig of 205157 bp in length.

FEATURES

Location/Qualifiers  
 1..205157  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-14C10"

BASE COUNT 53433 a 46205 c 47212 g 58163 t 144 others  
 ORIGIN

alignment\_scores:

Quality: 139.00 Length: 130  
 Ratio: 2.138 Gaps: 4  
 Percent Similarity: 50.000 Percent Identity: 32.308

alignment\_block:

US-09-060-609-2 x AC016672 ..

Align seg 1/1 to: AC016672 from: 1 to: 205157

```

138 TyrThrAlaHisValSerCysPheProAlaProAsn..... 149
|||||.....|.....|.....|.....|.....|.....|.....|
15567 TACTTTGGACATTTAAACTGCTCTCTGCCGTAATTTTGGTGTGTC 15616

150 .....1leThrCysLysAspSerSerGlyAsnGluThrHisP 162
|||||.....|.....|.....|.....|.....|.....|.....|
15617 AGTTGTGAAATAGTCACTGTGTTCCAGTTAAGGGCTACGATCAGGGG. 15665

162 heThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal 178
|||||.....|.....|.....|.....|.....|.....|.....|
15666 .....AAGCCCTTTCTCAATCTGCACATA 15689

179 Asn.....GlyTyrSerTyrLysVa 185
|||||.....|.....|.....|.....|.....|.....|.....|
15690 ATGTGTTGATTAAACCGTGACTGGCCACAGCTGGGTGGAGGTGCTCA 15739

185 lAlavalAla.....LeuSerLeuPheLeuGlyTrpLeuGlyA 198
|||||.....|.....|.....|.....|.....|.....|.....|
15740 TGGATCTCTGCTGTTTGGCCATCGCAGCATCACCTCGGTGGTTGGAG 15789

198 lAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLysPheCys 214
|||||.....|.....|.....|.....|.....|.....|.....|
15790 CAGACCGTTTCTACCTGGCCAGTCGGCGGAGGCTCGGCAAGCTCTTC 15839

215 ThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSe 231
::: |||.....|.....|.....|.....|.....|.....|.....|

```



15840 AGCTTCGTCGCTGGGATATGACGCTGATAGACGCTCGCTCATGG 15889

231 rMetGlnIleValGlyProSerAspGlySerTyrIle 244

15890 AGTTGGCTATGTTGGACGACGATGGCTCTTTGACATT 15929

seq\_name: gb\_inl:CELC02F5

seq\_documentation\_block:

LOCUS CELC02F5 22333 bp DNA INV 22-OCT-1993  
DEFINITION C. elegans cosmid C02F5.  
ACCESSION L14745 L18807  
VERSION L14745.1 GI:289607  
KEYWORDS GRR1 gene; GTP-binding protein; glucose repression; proteasome component C5.  
SOURCE Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed whole animal DNA.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 22333)  
Sullivan, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterston, R.  
The C. elegans genome sequencing project: A beginning  
Nature 356, 37-41 (1992)  
92168156

2 (bases 1 to 22333)  
Anderson, K.

Sequence of the C. elegans cosmid C02F5

Unpublished (1993)

Submitted by:

Nematode Sequencing Project

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

MRC Laboratory of Molecular Biology,

Cambridge CB2 2QH, England

e-mail: rwnematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk NOTE:

Coding sequences below are predicted from computer analysis, using

the program Genefinder (P. Green and L. Hillier, in preparation)

Neighboring cosmid information:

Bases 1 to 800 of this entry correspond to bases 26944 to 27743 of

GenBank entry CELC30A5 (L10990). Bases 22034 to 22333 of this

entry

correspond to bases 1 to 300 of GenBank CELF09G8 (L11247). The

sequence

of C. elegans cosmid C02F5 begins at position 18944 of GenBank

CELC30A5

(L10990), continues through this entry and ends at position 15818

of

GenBank CELF09G8 (L11247). The sequence of C. elegans cosmid F09G8

begins at position 22034 of this entry.

Location/Qualifiers

1. .22333

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/dev\_stage="mixed"

/sex="hermaphrodite"

/tissue\_type="whole animal"

/map="3"

join(8...124,175...257,305...694,742...943,993...1493,

2247...2423)

/standard\_name="C02F5.7 protein"

/note="homology with glucose induced repressor, GRR1;

putative"

/codon\_start=1

/protein\_id="AAA27922.1"

/db\_xref="GI:289614"

/translation="METAEIRISAAASRRRAKRLAQQAHKHPVIAQAKONMYLIT  
TLPQAQVDSLNRVLPKEVLLKVFSLDTKALCRSAOVCRSWILALDGSNNQRYDL  
FTFQRDVKAVENLARRCGGFLKSLKGCENVHDSALFTFSRCPNLEHLSLYRCK

exon

/note="putative"

/number=9

305...694

/note="putative"

/number=10

742...943

/note="putative"

/number=11

993...1493

/note="putative"

/number=12

2786

misc\_feature

/note="possible trans-spliced leader site for C02F5.6"

join(2793...2834,3022...3473,3528...3686,3737...4091,

4139...4252,4326...4463)

/standard\_name="C02F5.6 protein"

/note="putative"

/codon\_start=1

/protein\_id="AAA27921.1"

/db\_xref="GI:289613"

/translation="MAHTSDGHWGAPYDQNTYVAYEQLEIALLEPLDILETANVEEF

RPAFKHQDPNDRKNKNDDEWRDSIYNATDESDDHEQRKNFPQPPQVQVNSF

VKNTMEFRSSQIDISRLAVMGCGEMSEKGEYGLSGFTINVLSVIDEPLSISIG

QQLGKAESIFLRKHLERNAEILAVETGLPMLRYSVYGGDILEPHRFADVAIVSMEV

VEHIPLPNAKFFENVGLTLMPRIFIFSTPNHEYNVAFGMEPGFRRGHDKHFEWNRKE

FSNWLELSIRFPHYQIDPHYIGMTGTYENLSGASOAAVCRLOVDLNTTLPOEVTYPY

EMVGHLPCLRGSLIAYNLVKEAPLDWLEKLELQEHREPRDGCISFYIFNVQNLHLH

KAPVSFALTIDERKAMQVSSVIGTVEF"

3022...3473

/note="putative"

/number=13

3528...3686

/note="putative"

/number=14

3737...4091

/note="putative"

/number=15

4139...4252

/note="putative"

/number=16

join(6347...6436,6481...6795,6876...6992)

/standard\_name="C02F5.5 protein"

/note="putative"

/codon\_start=1

/protein\_id="AAA27920.1"

/db\_xref="GI:289612"

/translation="MSYNEQIPPEYKYVFGTRCAVLACSSFELMMVLLSLGSDNSV

LAKLVLVFIGASAAVSGENLQONIDGRDIKKTIGSDNETRGAALVIGIPALAAV

FVFLCVSGHAFPSFLVHLVHVLASIGQLGIEAYEVNRNGFCIKAPENKKNVADSYQNYO

ILNMVQDSVFDKK"

6481...6795

/note="putative"

/number=17

complement(join(8274...8477,8518...8583,8631...8897,

9041...9178,9473...9532))

/standard\_name="C02F5.8 protein"

/note="putative"

/codon\_start=1

/protein\_id="AAA27923.1"

/db\_xref="GI:289615"

/translation="MATWKFIIIRSVLFFLOLAMLALALIAVGFWMGYDSSFDLKL

NYLYKDPKSLADAKPFIIRWLIIVFWSIIGSLGAVVTAVLGNISWPKRGKFM

TVYLIIVLVSLGEGGAVLVRNLSHDNTNSLIDAMYTNSNDLKIIQDKVNC

IENSLNVMYCGPMQSKPHCDVAFDSDVNTMTISGILLVILITQITAILPVPILI

SRKTKYKSYEPRVTQLADITDTRF"

8518...8583

exon

```
/note="complement; putative"
/number=6
8631..8897
/note="complement; putative"
/number=7
9041..9178
/note="complement; putative"
/number=8
9558
/note="possible trans-spliced leader site for CO2F5.8"
complement(join(11079..11495,11809..12165,12215..12369,
13369..13474,13523..13916,13962..14227))
/standard_name="CO2F5.9 protein"
/note="homology with proteasome component C5; coded for by
C. elegans cDNAs GenBank:M80127 and T01498; putative"
/codon_start=1
/protein_id="AAA27924.1"
/db_xref="GI:289616"
/translation="MTSFTGITAVANATNEMAFKQAMKEVAHNPWSSRQIERQRW
NPYSMEGSCAIGENFAIVASDTRMTQNDINILTRDAEKIQLINDIILITISGFYG
DVQLKKVQLSRHKYRFDYRSDMSVDLCAELLSRLNYRFRFYITDGAIGADIEHG
KGAVFSYDPIERLIGYASGAAEMPIPELDCQIGHVTLSEGYRPELIDRAISL
MKDSFGAAERIEITGDKIHVIAEAGKPVVVKELQISISAQSONVLKPKFKISAE
LGFAPLNLYDTLGLKPRSDAMSGKRPASVLKGLDDWPTANLIPGSSIPSD
PSSALKVPKVLDESEILQITSGSSFSLNDSEQAQNEQERNIEDLLRQDMTESK
IKNEASLLGLKGLGELESSESHANYVITCNFFCGGLKNRKYKGLRLLDHALTH
SNRIPCKLACFECTNVRMRSHYAKAHNPNERVEGYGMKALVSGDSSRIGDVGDDID
QVSDDELKELNWCFSIESIHLVQASGFEVGEKYRRTKRRKLDREAMSNWFNF"
11809..12165
/note="complement; putative"
/number=2
12215..12369
/note="complement; putative"
/number=3
13369..13474
/note="complement; putative"
/number=4
13523..13916
/note="complement; putative"
/number=5
join(14580..14717,14761..14894,14937..15120,15330..15412,
15465..15612,15661..15759)
/standard_name="CO2F5.4 protein"
/note="coded for by C. elegans cDNA GenBank:CELL16D4
(M89109); putative"
/codon_start=1
/protein_id="AAA27919.1"
/db_xref="GI:289611"
/translation="MAFTEQTLRQKLTNLNHPSSSIQTTSALLQNSRELIIIRWV
LKTVKTHGSKIYNLVANDVSONARKTCPOFKDEFFPAIESFRHAIELKNAKEV
EHAIGKLTINVKDRIQFTPSQCKRLHEVQOVKLSGFSPTPAVANKHGKSVRSQEA
QSLVQNEEALPMLEEYVKRLKNETNERETLESNLMLENVRMSIEHHEKLCREVR
REDRIKADLLEVEFTFSLPLAEMPNAPIPTLEALFEKKK"
14761..14894
/note="putative"
/number=18
14937..15120
/note="putative"
/number=19
15330..15412
/note="putative"
/number=20
15465..15612
```

## alignment\_scores:

Quality: 138.50 Length: 78  
Ratio: 2.473 Gaps: 1  
Percent Similarity: 71.795 Percent Identity: 38.462

## alignment\_block:

US-09-060-609-2 x CELC02F5 ..

Align seg 1/1 to: CELC02F5 from: 1 to: 22333

```
179 AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
18579 AATGGTCACACTACTTCTCACAACACTCTCTCTGATTATCCATTTTCTCGGTG 18628
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

195 pLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuLeuLeuL 212
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
18629 GGTTGCTGTCGACGGTTTTCCTCGGATATCTGCAATGGCTGCGGAA 18678
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

212 ySPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
18679 AACTAATGACTTTGGGTGGATTGGAACTCGGTGATTCGATATATTC 18728
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSerTyrIleI 245
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
18729 TTACTGTGACTTTGGAGTACTCGGACCGCGGATGACTCGAGTTGG...GA 18775
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

245 eAspTyrTyrGlyThrArgLeuThrArgLeuSer 256
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
18776 GCGGTACTATTAAGAAAGAAATTAATAACGAGT 18809
|||||: : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: gb_htg6:AC011069

seq_documentation_block:
LOCUS AC011069 168266 bp DNA HTG 02-DEC-1999
Drosophila melanogaster chromosome X clone BACR11H20 (D881) RPCI-98
11.H.20 map 12B-12C strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
*** 92 unordered pieces.

ACCESSION AC011069
VERSION AC011069.6 GI:6503260
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 168266)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,
Hinkler,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Poshrefi,B.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Wooley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 168266)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
```

## Direct Submission

Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Dec 2, 1999 this sequence version replaced gi:6492479.

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 92 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.  
\* 355: contig of 355 bp in length  
\* 356: gap of unknown length  
\* 435: contig of 323 bp in length  
\* 436: gap of unknown length  
\* 759: gap of unknown length  
\* 839: contig of 900 bp in length  
\* 1739: gap of unknown length  
\* 1819: contig of 981 bp in length  
\* 2800: gap of unknown length  
\* 2880: contig of 720 bp in length  
\* 3600: gap of unknown length  
\* 3680: contig of 649 bp in length  
\* 4329: gap of unknown length  
\* 4409: contig of 497 bp in length  
\* 4906: gap of unknown length  
\* 4986: contig of 641 bp in length  
\* 5627: gap of unknown length  
\* 5707: contig of 532 bp in length  
\* 6239: gap of unknown length  
\* 6319: contig of 628 bp in length  
\* 6947: gap of unknown length  
\* 7027: contig of 1123 bp in length  
\* 8150: gap of unknown length  
\* 8230: contig of 638 bp in length  
\* 8868: gap of unknown length  
\* 8948: contig of 789 bp in length  
\* 9737: gap of unknown length  
\* 9817: contig of 751 bp in length  
\* 10568: gap of unknown length  
\* 10648: contig of 900 bp in length  
\* 11548: gap of unknown length  
\* 11628: contig of 937 bp in length  
\* 12565: gap of unknown length  
\* 12645: contig of 734 bp in length  
\* 13379: gap of unknown length  
\* 13459: contig of 1158 bp in length  
\* 14617: gap of unknown length  
\* 14697: contig of 660 bp in length  
\* 15357: gap of unknown length  
\* 15436: contig of 757 bp in length  
\* 16193: gap of unknown length  
\* 16273: contig of 814 bp in length  
\* 17088: gap of unknown length  
\* 17168: contig of 1633 bp in length  
\* 18801: gap of unknown length  
\* 18881: contig of 982 bp in length  
\* 19863: gap of unknown length  
\* 19942: contig of 1380 bp in length  
\* 21323: gap of unknown length  
\* 21402: contig of 1249 bp in length  
\* 22652: gap of unknown length  
\* 22731: contig of 1063 bp in length  
\* 23794: gap of unknown length  
\* 23874: contig of 1922 bp in length  
\* 25796: gap of unknown length  
\* 25877: contig of 1431 bp in length  
\* 27307: gap of unknown length  
\* 27387: contig of 2208 bp in length  
\* 29595: gap of unknown length  
\* 29675: contig of 2478 bp in length  
\* 32153: gap of unknown length  
\* 32154: contig of 2052 bp in length  
\* 34286: gap of unknown length  
\* 34365: contig of 1503 bp in length  
\* 35868: gap of unknown length  
\* 35948: contig of 1661 bp in length  
\* 37609: gap of unknown length  
\* 37610: contig of 1904 bp in length  
\* 37690: gap of unknown length  
\* 39594: contig of 2420 bp in length  
\* 42093: gap of unknown length  
\* 42094: contig of 2234 bp in length  
\* 44174: gap of unknown length  
\* 44408: gap of unknown length  
\* 44488: contig of 2558 bp in length  
\* 47046: gap of unknown length  
\* 47126: contig of 2547 bp in length  
\* 49672: gap of unknown length  
\* 49752: gap of unknown length  
\* 51251: contig of 1499 bp in length  
\* 51331: gap of unknown length  
\* 52547: contig of 1216 bp in length  
\* 52627: gap of unknown length  
\* 54821: contig of 2194 bp in length  
\* 54901: gap of unknown length  
\* 56540: contig of 1639 bp in length  
\* 56620: gap of unknown length  
\* 59274: contig of 2654 bp in length  
\* 59354: gap of unknown length  
\* 62336: contig of 2982 bp in length  
\* 62416: gap of unknown length  
\* 66304: contig of 3888 bp in length  
\* 66384: gap of unknown length  
\* 68911: contig of 2527 bp in length  
\* 68991: gap of unknown length  
\* 71844: contig of 2853 bp in length  
\* 71924: gap of unknown length  
\* 74959: contig of 3035 bp in length  
\* 75039: gap of unknown length  
\* 77200: contig of 2161 bp in length  
\* 77280: gap of unknown length  
\* 80390: contig of 3110 bp in length  
\* 80470: gap of unknown length  
\* 84267: contig of 3797 bp in length  
\* 84347: gap of unknown length  
\* 89197: contig of 4850 bp in length  
\* 89277: gap of unknown length  
\* 96088: contig of 6811 bp in length  
\* 96168: gap of unknown length  
\* 102457: contig of 6289 bp in length  
\* 102537: gap of unknown length  
\* 107506: contig of 4969 bp in length  
\* 107586: gap of unknown length  
\* 113433: contig of 5847 bp in length  
\* 113513: gap of unknown length  
\* 120777: contig of 7264 bp in length  
\* 120857: gap of unknown length  
\* 134488: contig of 13631 bp in length  
\* 134568: gap of unknown length  
\* 146676: contig of 12108 bp in length  
\* 146756: gap of unknown length  
\* 147420: contig of 664 bp in length  
\* 147500: gap of unknown length  
\* 148009: contig of 509 bp in length  
\* 148039: gap of unknown length  
\* 148420: contig of 331 bp in length  
\* 148500: gap of unknown length  
\* 149336: contig of 836 bp in length  
\* 149416: gap of unknown length  
\* 149628: contig of 212 bp in length  
\* 149708: gap of unknown length  
\* 150288: contig of 560 bp in length  
\* 150348: gap of unknown length  
\* 150645: contig of 297 bp in length  
\* 150725: gap of unknown length  
\* 151437: contig of 712 bp in length  
\* 151517: gap of unknown length  
\* 153103: contig of 1586 bp in length  
\* 153183: gap of unknown length  
\* 153649: contig of 466 bp in length  
\* 153729: gap of unknown length  
\* 154264: contig of 535 bp in length  
\* 154344: gap of unknown length  
\* 154946: contig of 602 bp in length  
\* 155026: gap of unknown length  
\* 155641: contig of 615 bp in length  
\* 155721: gap of unknown length  
\* 156401: contig of 680 bp in length



```
CDS
1486..2628
/gene="BBU04"
/translation="similar to PID:1174341 PID:1174349 GB:AE001584
percent identity: 81.23; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07693.1"
/db_xref="GI:6382382"
/translation="MLISFVKESLPMHKIKTTNNPRCNKYNKQVKKLIVLSTICYLN
KTHKKYQKTYLYFENENLRNGQTSILRTMQKIYRLQREIKVITNYQRMGVNSG
TEIYLKPNKDCIHKINQHFKEKTRFQNRVNYFNKNSDSRMGSVQCESCSNKK
NNIKERINIERQYQVINYFNKCFSEKILSILLNVDKDTMIKIKTKRTDIK
AKNNIYFPKSCSKQEKLLCNTQKLEKSGNSEQLETKQIYENKYKPHF
IENHKYSDLSVIRKKLEKSTIERKKENSKQDYKNLRTNIFNLILQAKETNIELKP
IIEKVLNKKIEYKKVRIYYSSELLEIINGKHYSNLKFRKKS"
2668..3656
/gene="BBU05"
2868..3656
/gene="BBU05"
/translation="similar to GB:D26185 SP:P37522 PID:467381
PID:580906 GB:AL009126 percent identity: 55.98; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="plasmid partition protein, putative"
/protein_id="AAF07694.1"
/db_xref="GI:6382383"
/translation="MKNKPNILTIASLKGVGKSTLTILFSVLLKDLGKKLIVDMOP
QNSITFSNVHNVIEITYNTYMLKGVSNVCVKGINDFIIPAHPILENFNLEPI
DYKILNLEISAVIEFDYILLDTPPHRDLFNALNTDHIIPQVVERWSIES
FTILNMENFOIKKITYTSIVEMQFIKRNLTLDTERLLYKGYGVKGIHFSN
GIKVFNELIPESSSESYKKAETLHLNLYINHNKRYIFKN"
14633..15235
/gene="BBU06"
14633..15235
/gene="BBU06"
/translation="similar to GB:U03641 PID:458218 percent identity:
61.64; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07695.1"
/db_xref="GI:6382384"
/translation="MEDIYMSKKKKKFLFNDRIEHNLIIDNLSNLTNLDNLMVYKD
LKQKLNLSNEIDTKIORMKILYIEIKRKLKYDGFKSFSOYLKTFVVAQTQAFYL
KLYSKILEGSSIDTIKELGPNLKKHIFKKKSFDFNNNKKSLDNKNISIRIFMRDK
ELYEFVQDIIKRASYIFKELFQNRKRDVLTIVTKYDNRN"
15349..15813
/gene="BBU07"
15349..15813
/gene="BBU07"
/translation="similar to PID:1174341 PID:1174349 GB:AE001584
percent identity: 81.71; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07696.1"
/db_xref="GI:6382385"
/translation="MIEILTKIKTEIAKKNKHIFTKSCSKQEKLEKILCNQKE
LEKSGCSEQLTFNQKIYENKYKPHFIENHKYSDLSYIKRLEKSEIKENKPKQ
DYESKINIFHIFIEQLKKKEINIEIPLKPLVKEYLNNQKKIKYTKVFDTYTR"
15791..16084
/gene="BBU08"
15791..16084
/gene="BBU08"
/translation="similar to GB:U03641 PID:458218 percent identity:
61.64; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein; identified by Glimmer2;
putative"
/gene="BBU09"
16603..16800
/gene="BBU10"
16603..16800
/gene="BBU10"
/translation="similar to GB:U03641 PID:458212 percent identity:
66.22; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07697.1"
/db_xref="GI:6382386"
/translation="MGIFVQKECKTMSKAVDEIYCHSCGCTIKKDAEICISGVRNK
QTEYNKLVFLCLFLGYLGVHRFYVIGKIGTGLYFTFGFLYVGLIDILIRATNK
FECK"
16603..16800
/gene="BBU11"
16603..16800
/gene="BBU11"
/translation="similar to SP:Q10618 PID:1322425 GB:AL123456
percent identity: 53.61; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF07698.1"
/db_xref="GI:6382387"
/translation="MIKTPILISDQFKALKLKMGEIIVFTDTVVAIGANIYDEVA
VRMNLVKKKSINEPLMLYDTEIKELSSHVPRKSAKILMERKSPGLTYLVKKSIK
IPKFINNLDVTAIRIPKNTALNLRKLNPLVSSANLSKRPSTFSFMAKELNG
LVGGTIMPKKNSDNGIESTIIEFSDKDNVILIRPGAITKKMIKRYLGRVKNYAK
ETKIDLKMFYSRLRIPVHTFRGDNIRKYLTKVLTIRDTFKFYFFFLDKKNI
ILFDSLIEVAENLYKELVNAEKYDQVIVEIVENEELGVAINNRIVKASSNNFM"
17918..18365
/gene="BBU12"
17918..18365
/gene="BBU12"
/translation="This region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
GB:L31615 GB:L31418 PID:520783 PID:551742 percent
identity: 61.90; identified by sequence similarity;
putative"
alignment_scores:
Quality: 117.00 Length: 63
Ratio: 2.659 Gaps: 1
Percent Similarity: 69.841 Percent Identity: 42.857
alignment_block:
US-09-060-609-2 x AE001582/rev
Align seg 1/1 to reverse of: AE001582 from: 1 to: 18753
173 IleserCys.....AgAsnValasnGlyTyrSerTyrLysVa 185
|||||
16440 ATTTCTTGGCGGGGACAGAAATAAACCGAAACACATAAATTGT 16391
185 LAIaValaLaleSerLeuPheLeuGlyTrpLeuGlyAlaAspargPheT 202
::: ||| |||||||||::: |||::: |||||||
```









Percent Similarity: 45.000 Percent Identity: 24.231

alignment\_block:

US-09-060-609-2 x AC013791 ..

Align seg 1/1 to: AC013791 from: 1 to: 180077

```
39 LeuLeuProPheSerLeuProLeuLeu.....GlyGlyGlyGlyse 52
||||:||||| ||| ||| ||||| ||||| |||||
69387 CTCATGCCCGAGTCTCTCCCGCAGAGCCAGCAGCGGGGTGGGGTGA 69436
52 rGlySerGlyGluLysValSerValSerLysMetAla.....A 65
||||| ||||| ||||| ||||| ||||| ||||| |||||
69437 GGGGTCGGGAGTGTGTG...GTCAGGAGGCTGGGAAGCAGCAGGTACA 69483
65 laAlaTrpProSer.....GlyProSerAlaProGluAla 76
||||| ||||| ||||| ||||| ||||| ||||| |||||
69484 TGGCCTGGCCAGCAATCAGCTATAGAGGACCCCGCAGTTCACAGGCCA 69533
77 ValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrThrGlu 93
||||| ||||| ||||| ||||| ||||| ||||| |||||
69534 .....TGG.....ATGGGTACTGG 69547
93 yPro.....TrpGlyAla..... 97
||||| ||||| ||||| ||||| ||||| ||||| |||||
69548 TCCATAGCCTGTAGTAGCTGGGCCGACAGCAGGAGGTGACGAGCGGG 69597
98 .....ValAlaThrSerAlaGlyGlyGlu 105
||||| ||||| ||||| ||||| ||||| ||||| |||||
69598 CAAGCATGACCCGCTGAATTCACCTCTCCGAGATCAGCGGCGAGCAGTA 69647
106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTrpIleCysLysAs 122
||||| ||||| ||||| ||||| ||||| ||||| |||||
69648 GATTCTCACAGGACGAGACCTT..... 69671
122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrT 139
||||| ||||| ||||| ||||| ||||| ||||| |||||
69672 .....GTTGTGAACCTGTGCACAGAGG 69693
139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155
||||| ||||| ||||| ||||| ||||| ||||| |||||
69694 GATCAAGGTTGAGCACT...CCCATGAGAACTAATACAGATGATCTG 69740
156 SerGlyAsnGlu.....ThrHisPheThrGlyAsnGluValGlu 168
||||| ||||| ||||| ||||| ||||| ||||| |||||
69741 GGGTGAACAGTTTCATCTGCGCCACCCACATCCGTGGA...AAATTCG 69787
168 yPhePheLys.....ProIleSerC 175
||||| ||||| ||||| ||||| ||||| ||||| |||||
69788 CTTCCATAAACTGGTCCCTGGTGGGAAATCTTGGGGGCCACTGAGCT 69837
175 ys.....ArgAsnValAsnGlyTyr 181
||||| ||||| ||||| ||||| ||||| ||||| |||||
69838 GTAGGATGAAGAGGCTGAGATAAGGGGACACCCAGAGGAGTCTGGGGTTC 69887
182 SerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAl 198
||||| ||||| ||||| ||||| ||||| ||||| |||||
69888 ATGTGGTGGCAGTCTGCGTGGTCTGCGGTTCTGGGTTCTGGGTCGTCG 69937
198 aaSPArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLysPheCysT 215
||||| ||||| ||||| ||||| ||||| ||||| |||||
69938 GTCTGGTGCAGTCTGGGATTCGGGTGGGTGGCA...GTGCTGGTACGG 69984
215 hrValGlyPheCysGlyIle.....GlySerLeuIleAspPhe... 227
||||| ||||| ||||| ||||| ||||| ||||| |||||
69985 TTCTGGGTTCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 70034
228 ....IleLeuIleSerMetGlnIleValGly 236
||||| ||||| ||||| ||||| ||||| ||||| |||||
70035 TGGCTGGCAGTCTGCTGGTACGAGTCTGGGG 70064
```

seq\_name: gb\_htg3:AC010844

seq\_documentation\_block:

LOCUS AC010844 92509 bp DNA HTG 08-OCT-1999  
DEFINITION Drosophila melanogaster chromosome X clone BACR03F03 (D882) RPCI-98  
03.F.3 map 12C-12D strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 100 unordered pieces.

ACCESSION AC010844  
VERSION AC010844.4 GI:6016642  
HTG: HTGS\_PHASE1.  
SOURCE fruit fly.

ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 92509)  
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 92509)

AUTHORS

Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Direct Submission

JOURNAL

REFERENCE

AUTHORS

COMMENT

Submitted (24-SEP-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Oct 8, 1999 this sequence version replaced gi:6006066.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 100 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1  
\* 664: contig of 664 bp in length  
\* 665  
\* 744: gap of unknown length  
\* 745  
\* 1298: contig of 554 bp in length  
\* 1299  
\* 1378: gap of unknown length  
\* 1379  
\* 1387: contig of 609 bp in length  
\* 1388  
\* 2067: gap of unknown length  
\* 2068  
\* 2719: contig of 652 bp in length  
\* 2720  
\* 2799: gap of unknown length  
\* 2800  
\* 3932: contig of 1133 bp in length  
\* 3933  
\* 4012: gap of unknown length  
\* 4013  
\* 4712: contig of 700 bp in length  
\* 4792: gap of unknown length  
\* 4793  
\* 5546: contig of 754 bp in length  
\* 5547  
\* 5626: gap of unknown length  
\* 5627  
\* 6428: contig of 802 bp in length  
\* 6429  
\* 6508: gap of unknown length  
\* 7234: contig of 726 bp in length  
\* 7235  
\* 7314: gap of unknown length  
\* 7315  
\* 8210: contig of 896 bp in length  
\* 8211  
\* 8290: gap of unknown length  
\* 8957: contig of 667 bp in length

```
* 8958 9037: gap of unknown length
* 9038 9940: contig of 903 bp in length
* 9941 10020: gap of unknown length
* 10021 10867: contig of 847 bp in length
* 10868 10947: gap of unknown length
* 10948 11592: contig of 645 bp in length
* 11593 11672: gap of unknown length
* 11673 12300: contig of 628 bp in length
* 12301 12380: gap of unknown length
* 12381 12785: contig of 405 bp in length
* 12786 12865: gap of unknown length
* 12866 13713: contig of 848 bp in length
* 13714 13794: gap of unknown length
* 13794 14731: contig of 938 bp in length
* 14732 14811: gap of unknown length
* 14812 15574: contig of 763 bp in length
* 15575 15654: gap of unknown length
* 15655 16762: contig of 1108 bp in length
* 16763 16842: gap of unknown length
* 16843 17720: contig of 878 bp in length
* 17721 17800: gap of unknown length
* 17801 18710: contig of 910 bp in length
* 18711 18790: gap of unknown length
* 18791 19193: contig of 403 bp in length
* 19194 19272: gap of unknown length
* 19273 20195: contig of 922 bp in length
* 20196 20275: gap of unknown length
* 20276 21358: contig of 1083 bp in length
* 21359 21438: gap of unknown length
* 21439 22656: contig of 1218 bp in length
* 22657 22736: gap of unknown length
* 22737 23743: contig of 1007 bp in length
* 23744 23823: gap of unknown length
* 23824 25367: contig of 1544 bp in length
* 25368 25447: gap of unknown length
* 25448 26594: contig of 1147 bp in length
* 26595 26674: gap of unknown length
* 26675 27986: contig of 1312 bp in length
* 27987 28066: gap of unknown length
* 28067 29052: contig of 916 bp in length
* 29053 29242: gap of unknown length
* 29243 30004: contig of 862 bp in length
* 30005 30968: contig of 964 bp in length
* 30969 31048: gap of unknown length
* 31049 32547: contig of 1499 bp in length
* 32548 32627: gap of unknown length
* 32628 33808: contig of 1181 bp in length
* 33809 33888: gap of unknown length
* 33889 34986: contig of 1098 bp in length
* 34987 35066: gap of unknown length
* 35067 35803: contig of 737 bp in length
* 35804 35883: gap of unknown length
* 35884 36294: contig of 411 bp in length
* 36295 36374: gap of unknown length
* 36375 37470: contig of 1096 bp in length
* 37471 37550: gap of unknown length
* 37551 39563: contig of 2013 bp in length
* 39564 39643: gap of unknown length
* 39644 41051: contig of 1407 bp in length
* 41051 41131: gap of unknown length
* 41131 42389: contig of 1259 bp in length
* 42390 42469: gap of unknown length
* 42470 44106: contig of 1637 bp in length
* 44107 44186: gap of unknown length
* 44187 45257: contig of 1071 bp in length
* 45258 45337: gap of unknown length
* 45338 46808: contig of 1471 bp in length
* 46809 46888: gap of unknown length
* 46889 49104: contig of 2216 bp in length
* 49105 49184: gap of unknown length
* 49185 50451: contig of 1267 bp in length
* 50452 50531: gap of unknown length

* 50532 52592: contig of 2061 bp in length
* 52593 52672: gap of unknown length
* 52673 55002: contig of 2330 bp in length
* 55003 55082: gap of unknown length
* 55083 57023: contig of 1941 bp in length
* 57103 57103: gap of unknown length
* 57104 59408: contig of 2305 bp in length
* 59409 59488: gap of unknown length
* 59489 60015: contig of 527 bp in length
* 60016 60095: gap of unknown length
* 60096 60452: contig of 357 bp in length
* 60453 60532: gap of unknown length
* 60533 61086: contig of 554 bp in length
* 61087 61166: gap of unknown length
* 61167 62041: contig of 795 bp in length
* 62042 62577: contig of 536 bp in length
* 62578 62657: gap of unknown length
* 62658 63205: contig of 548 bp in length
* 63206 63285: gap of unknown length
* 63286 63847: gap of unknown length
* 63848 64432: contig of 585 bp in length
* 64433 64512: gap of unknown length
* 64513 65290: contig of 778 bp in length
* 65291 65370: gap of unknown length
* 65371 66186: contig of 816 bp in length
* 66187 66266: gap of unknown length
* 66267 66724: contig of 458 bp in length
* 66725 68804: gap of unknown length
* 68805 67422: contig of 618 bp in length
* 67423 67502: gap of unknown length
* 67503 68287: contig of 785 bp in length
* 68288 68367: gap of unknown length
* 68368 68994: contig of 627 bp in length
* 68995 69074: gap of unknown length
* 69075 69731: contig of 657 bp in length
* 69732 69811: gap of unknown length
* 69812 70509: contig of 698 bp in length
* 70510 71259: contig of 670 bp in length
* 71260 71339: gap of unknown length
* 71340 71975: contig of 636 bp in length
* 71976 72055: gap of unknown length
* 72056 72257: contig of 201 bp in length
* 72257 72336: gap of unknown length
* 72337 73237: contig of 903 bp in length
* 73240 73319: gap of unknown length
* 73320 73974: contig of 655 bp in length
* 73975 74054: gap of unknown length
* 74055 74743: contig of 689 bp in length
* 74744 74823: gap of unknown length
* 74824 75343: contig of 520 bp in length
* 75344 76021: contig of 598 bp in length

alignment_scores:
Quality: 110.50 Length: 91
Ratio: 1.973 Gaps: 1
Percent Similarity: 61.538 Percent Identity: 29.670

alignment_block:
US-09-060-609-2 x AC010844 ..
Align seg 1/1 to: AC010844 from: 1 to: 92509
153 LysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPh 169
|||||:|||||:|||||:|||||:|||||:
2821 AAAGAGTCTAGGTAAACGACCTTTCGCGCCGCGGA..... 2859
169 ePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValA 186
|||:||||:||||:||||:||||:
2860 .....GTGCCCTGTGTCGCTACACGGATCCTACTTTGTGACCA 2899
```

```
186 laValAlaLeuSerLeuPheLeuGlyTyrLeuGlyAlaAspArgPheTyr 202
    :::: |||:::|||||:::||||| ||| |||
2900 CACTGATCTACAGCATGCTGGGTTCTCGGTATGGATCGCTTCTGT 2949
    :::: ||| ||| :::: |||::: |||::: |||
. 203 LeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCy 219
    ||||| :::: |||::: |||::: |||::: |||
2950 CTCGGTCAACGGGCAGGCTGTGGCAAACTGCTTACCATGGCGGCGT 2999
    |||::: |||::: |||::: |||::: |||::: |||
219 sGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleValG 236
    |||::: |||::: |||::: |||::: |||::: |||
3000 GGGCGTTTGGTGCATCATCGACGTCATCCTCCTCATCACCAACAATTGC 3049
    |||::: |||::: |||::: |||::: |||::: |||
236 lyProSerAspGlySerSerTyr 243
    |||::: |||::: |||::: |||::: |||::: |||
3050 TGCCCGAGGACGCGCAATTGG 3072
```

OM of: US-09-060-609-2 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Feb 11, 2000 8:09 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2.1/USPTO.spool/US09060609/runat\_09022000\_150439\_25205/app\_query.fasta.1  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.030 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US09060609 -NCPU=6 -ICPU=3 -NO\_XLPXY -WAIT  
-THREADS=1

Search information block:

Query: US-09-060-609-2  
Query length: 269  
Database: N\_Geneseq\_36:\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 50.600000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X05735	+ 1439.00	2529.46	2.7e-133	810	Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.
N_Geneseq_36:X41191	+ 779.00	1368.07	1.3e-68	440	Human secreted protein 5' EST S
N_Geneseq_36:X41259	+ 760.50	1334.97	9.3e-67	435	Human secreted protein 5' EST S
N_Geneseq_36:V40276	+ 108.00	163.53	0.1634	2162	Human equilibrative nucleoside
N_Geneseq_36:V49582	+ 108.00	165.98	0.1207	1721	Human fibrosarcoma cell line H
N_Geneseq_36:V49583	+ 108.00	170.61	0.0667	1119	Human fibrosarcoma cell line H
N_Geneseq_36:X20370	+ 105.00	170.12	0.0710	715	Borrelia burgdorferi polynucleo
N_Geneseq_36:V84557	+ 96.00	142.47	2.46	2127	Human secreted protein gene 14
N_Geneseq_36:T39751	- 91.50	137.99	4.38	1339	Human mucosal addressin cell a
N_Geneseq_36:X35280	- 91.50	137.99	4.38	1339	cDNA encoding human MADCAM-1.
N_Geneseq_36:X35291	- 88.00	127.03	17.85	2397	Human Glyt-2 transporter G1831
N_Geneseq_36:V22914	+ 87.00	125.26	22.40	2397	Human Glyt-2 transporter T836A
N_Geneseq_36:V22906	+ 87.00	125.26	22.40	2397	Human Glyt-2 variant gene sequ
N_Geneseq_36:V22900	+ 87.00	129.63	12.79	1397	Fragment of the human Glyt-2 g
N_Geneseq_36:T39750	+ 86.50	128.56	14.67	1624	Human mucosal addressin cell a
N_Geneseq_36:V80679	+ 86.50	132.06	9.36	1173	Human DNAX toll-like receptor
N_Geneseq_36:V72100	- 86.50	117.50	60.56	4541	Human RON receptor cDNA. Treat
N_Geneseq_36:X35279	- 86.50	128.56	14.67	1624	cDNA encoding human MADCAM-1.
N_Geneseq_36:N50491	- 86.00	123.97	26.44	2397	Sequence of herpes simplex vir
N_Geneseq_36:V22917	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter G1371
N_Geneseq_36:V22919	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter G1829
N_Geneseq_36:V22920	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter A1831
N_Geneseq_36:V22921	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter G2103
N_Geneseq_36:V22922	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter T2198
N_Geneseq_36:V22923	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter A2203
N_Geneseq_36:V22924	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter C342G
N_Geneseq_36:V22925	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter C352T
N_Geneseq_36:V22926	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter T7330
N_Geneseq_36:V22927	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter A773G
N_Geneseq_36:V22929	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter G951A
N_Geneseq_36:V22910	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter T6C m
N_Geneseq_36:V22911	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter G304A
N_Geneseq_36:V22912	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter C371T
N_Geneseq_36:V22913	+ 86.00	123.49	28.11	2397	Human Glyt-2 transporter C571T
N_Geneseq_36:V22905	+ 86.00	123.49	28.11	2397	Human Glyt-2 gene sequence. Hu
N_Geneseq_36:V22909	+ 86.00	123.49	28.11	2397	DNA encoding a human Glyt-2 tr
N_Geneseq_36:V22932	+ 86.00	123.49	28.11	2397	DNA encoding a human Glyt-2 tr
N_Geneseq_36:X24307	- 85.50	126.32	19.54	1696	Rat bone mineralisation protei
N_Geneseq_36:T42080	+ 85.00	128.82	14.18	1238	Rat LAMP coding sequence. DNA
N_Geneseq_36:T42081	+ 85.00	130.97	10.77	1014	Rat LAMP coding sequence. DNA
N_Geneseq_36:T42083	+ 85.00	131.90	9.56	930	Rat mature LAMP coding sequence

N\_Geneseq\_36:V22916 + 85.00 121.72 35.27 2397 Human Glyt-2 transporter A1  
N\_Geneseq\_36:V62176 - 84.50 79.01 8.4e+03 117213 HSV-2 strain SB5 Contig 1  
N\_Geneseq\_36:X04125 + 84.50 120.46 41.47 2483 Human Factor B encoding cDN  
N\_Geneseq\_36:Q20349 + 84.00 120.93 39.05 2189 Thromboplastin clone 2b-Apr

seq\_name: N\_Geneseq\_36:X05735

seq\_documentation\_block:

ID X05735 standard; mRNA; 810 BP.  
AC X05735;  
DT 27-APR-1999 (first entry)  
DE Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.  
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;  
human; Alzheimer's disease; ss.  
OS Homo sapiens.  
FH Key  
FT CDS  
FT 1..810  
FT /\*tag= a "BBP"  
FT /product= "BBP"  
PN W09846636-A2.  
PD 22-OCT-1998.  
PF 14-APR-1998: U07462.  
PR 16-APR-1997: US-064583.  
PA (AMHP) AMERICAN HOME PROD CORP.  
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,  
PI Walker SG;  
DR WPI: 99-080736/07.  
DR P-PSDB: W94291.  
PT Polynucleotide encoding beta-amyloid peptide binding protein - used  
PT to identify inhibitors of beta-amyloid peptide for treating  
PT Alzheimer's disease  
PS Claim 1; Pages 43-44; 59pp; English.  
CC This represents a nucleotide sequence encoding a beta-amyloid peptide  
CC binding protein (BBP). The polynucleotide comprising the entire BBP  
CC nucleotide sequence of clone Bp1-fl is deposited under the accession  
CC number ATCC 98617. The polynucleotide comprising a fragment of BBP  
CC (nucleotides 202-807 of the full length BBP) of clone pBK196 is deposited  
CC as ATCC 98399. Host cells transformed with a vector comprising the BBP  
CC nucleic acid are used for the recombinant production of the protein. The  
CC protein can be used in a method for diagnosing a disease characterised by  
CC aberrant expression of human beta-amyloid protein (BAP). The protein can  
CC also be used in a method for screening for compounds which regulate  
CC expression of a BAP binding protein. The proteins, antibodies and  
CC identified compounds can be used in the treatment or prevention of  
CC Alzheimer's disease.  
SQ Sequence 810 BP; 204 A; 183 C; 202 G; 221 T;

alignment\_scores:

Quality: 1439.00 Length: 269  
Ratio: 5.349 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-060-2 x X05735

Align seg 1/1 to: X05735 from: 1 to: 810

1 MethistleLeuLysGlySerProAsnValIleProArgAlaHisGlyCl 17  
|||||  
1 ATGCATATTTAAAGGCTCTCCCAATGATTCCACGGGCTCAGGGCA 50  
|||||  
17 nLysAsnThrArgArgaspGlyThrGlyLeuTyProMetArgGlyProp 34  
|||||  
51 GAAGAACACGCGAGAGACGGAACTGCCCTCTATCTATGCGAGGTCCT 100  
|||||  
34 heLysAsnLeuAlaLeuLeuProPheSerLeuProLeuLeuGlyGly 50  
|||||  
101 TTAAGAACCTCGCCCTGTGCGCTTCCTCCCTCCCTCGCTCGGCGAGGC 150  
|||||  
51 GlySerGlySerGlyGlyLuLysValSerValSerLysMetAlaAlaTr 67  
|||||  
151 GGAAGCGGAAGTGGCGAAGAGTCTCGGTCTCCAAAGATGGCGCGCCTG 200  
|||||

67 pProSerGlyProSerAlaProGluAlaValThrAlaArgLeuValGlyV 84  
 201 GCCCTCTGTCGTCGTCGAGGAGCGTGACGCCAGCTCGTTGGTG 250  
 84 alLeuTrpPheValSerValThrThrGlyProTrpGlyAlaValAlaThr 100  
 251 TCCCTGTGTTGCTCTAGCTACTACAGACCTCGGGGGCTTTGCCACC 300  
 101 SerAlaGlyGlyGluSerLeuLysCysGluAspLeuLysValGlyG1 117  
 301 TCCGCCGGGGGAGGAGTCGCTTAAGTGGAGGACCTCAAAAGTGGACA 350  
 117 nTyrlleCysLysAspProLysIleAsnAspAlaThrGlnGluProValA 134  
 351 ATATATTGTTAAAGATCCAAAATAAATGACGCTACGGAAGAACAGTTA 400  
 134 snCysThrAsnTyThrAlaHisValSerCysPheProAlaProAsnIle 150  
 401 ACTGTACAAACTACACAGCTATGTTCTCTGTTTCCAGCACCCCAACATA 450  
 151 ThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluVa 167  
 451 ACTGTAGGATTCAGTGGCAATGAACACATTTTACTGGGAAGCAAGT 500  
 167 lGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrlSerTyrl 184  
 501 TGGTTTTTCAAGCCCATATCTGCGCAATGTAATGGCTATTCCTACA 550  
 184 yValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArg 200  
 551 AAGTGGCAGTCGCATGTCTCTTTTCTGGATGTTGGGAGCAGATCGA 600  
 201 PheTyrlleuGlyTyrlProAlaLeuGlyLeuLeuLysPheCysThrValG1 217  
 601 TTTTACCTGGATACCTCGCTTTGGTTGTTAAAGTTTGCACCTGTAGG 650  
 217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234  
 651 GTTGTGGAAATGGGACCCCAATTGATTTCATCTTATTTCATTCAGCA 700  
 234 leValGlyProSerAspGlySerSerTyrlleIleAspTyrlTyrlThr 250  
 701 TTGTGGACCTTCAGATGGAAGTAGTACATATATAGATTACTATGGAACC 750  
 251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnIle 267  
 751 AGACTTAAAGACTGAGTATTACTAATGAACATTTAGAAAAACGAATT 800  
 267 uTyrlPro 269  
 801 ATATCCA 807

seq\_name: N\_Geneseq\_36:X41191

seq\_documentation\_block:

ID X41191 standard; cDNA; 440 BP.

DE 17-JUN-1999 (first entry)

KW Human secreted protein 5' EST SEQ ID NO:135.

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

PN WO9006548-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; IBI222.

PR 01-AUG-1997; US-905135.

PA (GSE ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI; 99-153778/13.  
 DR P-PSDB; Y12358.  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 PS Claim 1; Page 315; 824pp; English.  
 CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SQ Sequence 440 BP; 107 A; 103 C; 114 G; 114 T;

alignment\_scores:

Quality: 779.00 Length: 146  
 Ratio: 5.372 Gaps: 0

Percent Similarity: 99.315 Percent Identity: 99.315

alignment\_block:

US-09-060-609-2 x X41191

Align seg 1/1 to: X41191 from: 1 to: 440

56 GluLysValSerValSerLysMetAlaAlaTrpProSerGlyProse 72  
 3 GAGAAAGTGTGCTCCAAAGATGGCGCGCTGCTGCTGCTCGTC 52  
 72 rAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrpPheValS 89  
 53 TGCTCGGAGCGCGTACGCGCCAGACTCGTTGGTGTCTGCTGCTGCT 102  
 89 exValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyGlyGlu 105  
 103 CAGTCACTACAGACCTCGGGGGCTGTGGCCACCTCCCGCGGGGGCGAG 152  
 106 GluSerLeuLysCysGluAspLeuLysValGlyGlnTyrlleCysLys 122  
 153 GAGTGCCTTAAGTGGCAGGACCTCAAAGTGGCAATATATTGTAAGA 202  
 122 pProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsnTyrt 139  
 203 TCCAAAAATAAATGACGCTACGCAAGAACAGTAACTGTACAAACTACA 252  
 139 hrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAspSer 155  
 253 CAGCTCATGTTTCTCTGTTTTCAGCACCCCAACATAACTTGTAGGATTC 302  
 156 SerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheLysPr 172  
 303 AGTGGCAATGAACACATTTTACTGGGAACGAAGTTGGTTTTTCAAGCC 352  
 172 oLleSerCysArgAsnValAsnGlyTyrlSerTyrlLysValAlaValAla 189  
 353 CATATCTTCCCGAAATGTAATGGCTATTCTTACAAAGTGGCAGTCGCAT 402  
 189 euSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPhe 201  
 403 TGCTCTCTTTTCTTGGATGTTGGGAGCAGATCGATT 440

seq\_name: N\_Geneseq\_36:X41259

## seq\_documentation\_block:

ID X41259 standard; cDNA; 455 BP.  
 AC X41259;  
 DE 17-JUN-1999 (first entry)  
 DT Human secreted protein 5; EST SEQ ID NO:203.  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 OS Homo sapiens.  
 PN W0906548-A2.  
 PD 11-FEB-1999.  
 PF 31-JUL-1998; IB1222.  
 PR 01-AUG-1997; US-905135.  
 PA (GEST) GENSET.  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 P-PSDB; Y12426.  
 FT New nucleic acids encoding human secreted proteins - obtained from  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 PS Claim 1; Page 456; 824pp; English.  
 CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SQ Sequence 455 BP; 102 A; 107 C; 115 G; 122 T;

## alignment\_scores:

Quality: 760.50 Length: 152  
 Ratio: 5.173 Gaps: 2  
 Percent Similarity: 96.711 Percent Identity: 94.079

## alignment\_block:

US-09-060-609-2 x X41259 ..

Align seg 1/1 to: X41259 from: 1 to: 455

60 ValSerLysMetAlaAlaLaTrpProSerGlyProSerAlaProGluAl 76  
 |||||  
 3 GTCTCAAGATGGCGCGCGCTGGCGCTCGGCTCGCTCGGAGG 52  
 |||||  
 76 aValThrAlaArgLeuValGlyValLeuTrpPheValSerValThrG 93  
 |||||  
 53 CGTGAGCGGCAGACTCGTGGTCTCTGTGGTCTGCTCAGTCACTACAG 102  
 |||||  
 93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLeuLys 109  
 |||||  
 103 GACCCCTGGGGGCTGTGCCACCTCCGCCGGGCGAGGAGTGCCTAAG 152  
 |||||  
 110 CysGluAspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAs 126  
 |||||  
 153 TCGGAGGACCTCAAGTGGGACAAATATATTGTAAAGATCAAAATAAA 202  
 |||||  
 126 nAspAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisVal 143  
 |||||

203 TGAGCGCTACGCAAGAACCCAGTTAACTGCTACAACTACACAGCTCATGTTT 252  
 143 erCysPheProAlaProAsnIleThrCysLysAspSerSerGlyAsnGlu 159  
 |||||  
 253 CCGTGTTCCTCCAGCACCCACACATACCTTGTAGGATNCCAGTGGCAATGAA 302  
 160 ThrHisPheThrGlyAsnGluValGlyPheLysProIleSerCysAr 176  
 |||||  
 303 ACACATTTTACTGGGAAGAGTGGTTTTTTTCAAGCCCATATCTTGGC 352  
 176 gAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuphe 192  
 |||||  
 353 AAATGTAATGGGTATTCCTACNNTRKAGCAGT...NNNWTGCTCTCTTTT 399  
 193 LeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuG1 209  
 |||||  
 400 CTGGATGGTGGGAGCAGATCGATTTTACCTGGATACCTGCTTGGG 449  
 209 yLeu 210  
 |||||  
 450 TTG 453

seq.name: N.Geneseq\_36:V40276

## seq\_documentation\_block:

ID V40276 standard; cDNA; 2162 BP.  
 AC V40276;  
 DT 13-OCT-1998 (first entry)  
 DE Human equilibrative nucleoside transporter 1 encoding cDNA.  
 KW Human; equilibrative nucleoside transporter; hENT1; hENT2; hENT1;  
 KW hENT2; coronary; cerebrovascular anoxia; viral infection; cancer; ss;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 179..1549  
 FT /\*tag= a  
 FT /product= "equilibrative nucleoside transporter 1"

WO9829437-A2.

PD 09-JUL-1998.

PF 30-DEC-1997; IB1657.

PR 03-NOV-1997; US-064004.

PR 30-DEC-1996; US-034083.

PA (UYAL-) UNIV ALBERTA.

PA (UYLE-) UNIV LEEDS.

PI Baldwin SA, Cass CE, Young JD;

DR WPI; 98-388035/33.

DR P-PSDB; W69555.

PT Newly isolated equilibrative nucleoside transporter protein(s) and

PT gene(s) - used to develop products for treating disorder(s)

PT associated with the transporter(s) and for use with nucleoside

PT drug(s)

PS Claim 21; Fig 2; 97pp; English.

CC The present sequence encodes a substantially purified equilibrative  
 CC nucleoside transporter (ENT), human ENT1 (hENT1). ENTs can transport a  
 CC variety of purines and pyrimidines, including adenosine, uridine,  
 CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are  
 CC bidirectional, they transport a suitable permeant both into and out of  
 CC cells. ENTs can be used as a tool for the development of new nucleoside  
 CC drugs. Products from the present invention can be used for treating a  
 CC subject having a disorder associated with an ENT. They can also be used  
 CC with nucleoside drugs in the treatment of e.g. coronary or  
 CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.  
 CC antibodies and oligonucleotides hybridising to nucleic acid sequences  
 CC encoding ENTs) can also be used for detection and drug screening.  
 SQ Sequence 2162 BP; 389 A; 638 C; 560 G; 575 T;

## alignment\_scores:

Quality: 108.00 Length: 236  
 Ratio: 0.947 Gaps: 14  
 Percent Similarity: 48.305 Percent Identity: 23.305

## alignment\_block:

US-09-060-609-2 x V40276 ..







[illegible]





143 erCysPheProAlaProAsnIleThrCysLysAspSerSergly 157  
 729 GGTGTGGGAGACTCCGGGAGGTGGTGTGGGAGGCTCCGGGG 686

seq\_name: N\_Geneseq\_36:V22918

seq\_documentation\_block:

ID\_V22918 standard; CDNA; 2397 BP.

AC V22918;

DT 29-JUL-1998 (first entry)

DE Human Glyt-2 transporter G1387A mutant gene sequence.

KW Human; Glyt-2 transporter; glycine transporter; neurotransmitter;

KW activity; N-methyl-D-aspartate receptor; NMDA;

KW strychnine-sensitive glycine receptor; treatment; diagnosis;

KW nervous system disorder; myoclonus; muscle spasm; hyperactivity;

KW epilepsy; spasticity; head trauma; neuronal cell death;

KW multiple sclerosis; spinal cord injury; Huntington's disease;

KW amytrophic lateral sclerosis; ds.

OS Synthetic.

MO Homo sapiens.

Key Location/Qualifiers

mutation 1387

FT /tag= a

FT /note= "G1387 replaced with A"

PN WO9807854-Al.

PD 26-FEB-1998.

PF 20-AUG-1997; U14637.

PR 20-AUG-1996; US-700013.

PA (ALLX) ALLELIX NEUROSCIENCE INC.

PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;

DR WPI; 98-169138/15.

PT Human glycine transporter - useful for treating nervous system

PT disorders, e.g. pain, myoclonus, etc.

PS Claim 6; Page -; 90pp; English.

CC The present sequence represents a human Glyt-2 transporter mutant

CC gene sequence. Glyt-2 is a glycine transporter found predominantly in

CC the brain stem and the spinal cord. Glyt-2 regulates the levels of

CC glycine, a major neurotransmitter, that preferentially influences the

CC activity of N-methyl-D-aspartate (NMDA) receptors and

CC strychnine-sensitive glycine receptors. The Glyt-2 protein can be

CC used to treat or to diagnose a nervous system disorder or condition,

CC e.g. pain, myoclonus, muscle spasm or hyperactivity, epilepsy or

CC spasticity associated with stroke, head trauma, neuronal cell death,

CC multiple sclerosis, spinal cord injury, dystonia, Huntington's disease

CC or amytrophic lateral sclerosis.

CC note: the present sequence does not appear in the specification; it was

CC created using information provided.

Sequence 2397 BP; 509 A; 665 C; 628 G; 595 T;

alignment\_scores:

Quality: 88.00 Length: 302

Ratio: 0.624 Gaps: 17

Percent Similarity: 46.689 Percent Identity: 22.517

alignment\_block:

US-09-060-2 x V22918

Align seg 1/1 to: V22918 from: 1 to: 2397

25 ThrGlyLeuTrpMetArgGlyProPheLysAsnLeuAla.....Le 39

|||||.....|

591 ACTGGACTTCATCTGTCATGTTGGGTGGGTACGAGTGGGCTGGCAATG 640

|||||.....|

39 uLeuProPheSerLeuLeuGlyGlyGly..... 51

|||||.....|

641 TCTGGAGGTTCCCTACCTGGCCCTCCAGAACGGGGAGGTCTTCTC 690

|||||.....|

52 .....SergLysSerGly.Glu..... 56

|||||.....|

691 ATCCCTTACCTGATGATGCTGGTCTGGCTGGATTACCCATCTTCTT 740

|||||.....|

57 .LysValSerValSerLysMetAlaAlaTrpPro..... 68

|||||.....|

741 GGAGGTGTCGTGGCCAGTTTCCAGCCAGGACAGTGTGTGTGGG 790

|||||.....|

69 .....SergLysProSerAlaProGluAlaValThrAla 79

|||||.....|

791 AGGCCATCCAGCTCTACAAGGCTGTGGCATCGGATGTCATCTCT 840

|||||.....|

80 ArgLeuValGlyValLeuTrp..... 86

|||||.....|

841 GTCCTAATAGCATATACATACTACAATGTGATTATTGCTATACACTTTCTA 890

|||||.....|

87 .....PheValSerValThrThrGlyProTrpGlyAlaVala 99

|||||.....|

891 CCTGTTTCCCTCTTGTGTGTCTACTA.....CCTGGGGCTCCTGCA 934

|||||.....|

99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp..... 112

|||||.....|

935 ACAACCTTGGATACGCCAGAA.....TGCAAGATAAACCAA 975

|||||.....|

113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 138

|||||.....|

976 CTTTATTAGATTCTCTGTGTATTAGTGACCATCCCAATA..... 1017

|||||.....|

128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145

|||||.....|

1018 .....CAGATCAAGAACTCGACTTTCTGCTGACCGCTT 1051

|||||.....|

145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161

|||||.....|

1052 ATCCCAACGTGACAATGGTTAATTCACGACCGCCAGCCCAATAGACATTT 1101

|||||.....|

162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178

|||||.....|

1102 GTCAGTGGAGTGAA...GAGTACTTCAAGTACTTTGTCTGAGATTTC 1148

|||||.....|

178 lasnGlyTyrSerTyr.....LysValAlaValAlaLeuSerL 191

|||||.....|

1149 TGCAGGGATTGAATTCCTCGGCGAGATCAGGTGGCCACTAGCTCTCTGCC 1198

|||||.....|

191 euPheLeuGlyTyrLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207

|||||.....|

1199 TCTTCTGGTGGTGGTGC.....ATTGTGTATGATCG 1230

|||||.....|

208 Leu.....GlyLeuLeuLysPheCysThrValG 217

|||||.....|

1231 TTGGCTAAGGAATCAAGACTTCAGAAAGTGGTGTACTTCACGGCCAC 1280

|||||.....|

217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234

|||||.....|

1281 GTTCCCGTATGTC...GTACTCGTGTATCTCTCATCCGAGGAGTCACCC 1327

|||||.....|

234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250

|||||.....|

1328 TGCCTGGAGCTGGAGTGGGATCTGTGTACTTCTATCATCACCCAAAGTGGGAG 1377

|||||.....|

251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267

|||||.....|

1378 AAACTCACGAATGCCGCTGGAAAGATGCT.....GCCACTCAGAT 1421

|||||.....|

267 uTyr 268

|||||.....|

1422 TTTC 1425

seq\_name: N\_Geneseq\_36:V22918

seq\_documentation\_block:

ID\_V22914 standard; CDNA; 2397 BP.

AC V22914;

DT 29-JUL-1998 (first entry)

DE Human Glyt-2 transporter T836A mutant gene sequence.

KW Human; Glyt-2 transporter; glycine transporter; neurotransmitter;

KW activity; N-methyl-D-aspartate receptor; NMDA;

KW strychnine-sensitive glycine receptor; treatment; diagnosis;



```
PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR WPI: 98-169158/15.
DR P-PSDB: W56369.
PT Human glycine transporter - useful for treating nervous system
PS disorders, e.g. pain, myoclonus, etc.
PS Example 1A; Pages 47-48; 90pp; English.
CC The present sequence encodes a variant of the human Glyt-2 transporter.
CC The cDNA encoding Glyt-2 was generated by reverse transcription
CC PCR (RT-PCR) from human spinal cord or cerebellum polyA mRNA. Glyt-2 is
CC a glycine transporter found predominantly in the brain stem and the
CC spinal cord. Glyt-2 regulates the levels of glycine, a major
CC neurotransmitter, that preferentially influences the activity of
CC N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine
CC receptors. The glyt-2 protein can be used to treat or to diagnose a
CC nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm
CC or hyperactivity, epilepsy or spasticity associated with stroke, head
CC trauma, neuronal cell death, multiple sclerosis, spinal cord injury,
CC dystonia, Huntington's disease or amyotrophic lateral sclerosis.
SQ Sequence 2397 BP; 509 A; 662 C; 629 G; 597 T;

alignment_scores:
Quality: 87.00 Length: 302
Ratio: 0.617 Gaps: 17
Percent Similarity: 46.689 Percent Identity: 22.517

alignment_block:
US-09-060-609-2 x V22906 ..
Align seg 1/1 to: V22906 from: 1 to: 2397

25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla.....Le 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
591 ACTGGACATTCATCCTCCATCCAGTGGGGTACGACGAGTGGGGCTGGCAATG 640

39 uLeuProPheSerLeuProLeuLeuGlyGlyGly..... 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 TCTGGAGGTTCCCTACCTACCGCTTCCAGAACGGGGAGGTGCTTTCTCTC 690

52 .....SerGlySerGly.Glu..... 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
691 ATCCCTTACCTGATGCTGCTGGCTGCTGGGTGATGATACCATCTCTTCTT 740

57 LysValSerValSerLysMetAlaAlaLatrpPro..... 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
741 GGAGGTGCTGGCGCCAGTTTCCAGCCAGGACGACGAGTCTGTGTGGA 790

69 .....SerGlyProSerAlaProGluAlaValThAla 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
791 AGGCCATCCAGCTCTACAAGGCTGTGGCATCGCGATGCTGATCAACTCT 840

80 ArgLeuValGlyValLeuTrp..... 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841 GTCTTAATAGCCATATACATAAGTATGATATTGCTATACATTTTCTA 890

87 .....PheValSerValThrGlyProTrpGlyAlaVala 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 CCGTTTGGCTCCTTTGTCTCTACTA.....CCCTGGGGCTCTGTGCA 934

99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluasp..... 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
935 ACAACCCCTTGAATGACGCCAGAA.....TGCAAGATAAAACCAAA 975

113 LeuLysValGlyGlnTyrIleCysLysasp....ProLysIleAsnAspAl 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
976 CTTTATTATGATCTCTGCTGCTTATCATGACCAATCCCAATA..... 1017

128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysp 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 .....CAGATCAAGAACTCGACTTCTTCGATGACCGCTT 1051

145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
1052 ATCCCAACGTCGACAATGGTTAAATTTTACCAGCCAGCGCAATAAGACATTT 1101
162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1102 GTCAGTGAAGTGAAG...GAGTACTTCAAGTACTTGTGCTGAAGATTC 1148
178 lAsnGlyTyrSerTyr.....LysValAlaValAlaLeuSerL 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1149 TGCAGGAGTGAATATATCTGCGAGATCAGTGGCCACTAGCTCTCTGCC 1198
191 euPheLeuGlyTyrLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1199 TCTCTCTGGCTGGGTC.....ATTGCTATGTCATCG 1230
208 Leu.....GlyLeuLeuLysPheCysThrValG 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1231 TTGGCTAAAGGAATCAAGACTTTCAGGAAAGTGGTGTACTTACCGGCCAC 1280
217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnI 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1281 GTTCCCGTATGTC...GTACTCGTATCTCTCTCATCCGAGGAGTCACCC 1327
234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1328 TGCTGGAGCTGGAGTGGGATCTGGTACTTTCATCACACCCAAAGTGGGAG 1377
251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1378 AAATCTACGAGTGCACGCGTGTGAAAGATGCT.....GCCACTCAGAT 1421
267 uTyr 268
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1422 TTTC 1425

seq name: N_Geneseq_36:V22900
seq_documentation_block:
ID V22900 standard; CDNA; 1597 BP.
AC V22900;
DE 29-JUL-1998 (first entry)
DE Fragment of the human Glyt-2 gene sequence in clone phg2-2.
KW Human; Glyt-2 transporter; glycine transporter; neurotransmitter;
KW activity; N-methyl-D-aspartate receptor; NMDA;
KW strychnine-sensitive glycine receptor; treatment; diagnosis;
KW nervous system disorder; myoclonus; muscle spasm; hyperactivity;
KW epilepsy; spasticity; head trauma; neuronal cell death;
KW multiple sclerosis; spinal cord injury; Huntington's disease;
KW amyotrophic lateral sclerosis; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 2..535
FT /*tag= a
FT 536..1597.
FT /*tag= b
FT /*note= "no stop codon given"
PN WO9807854-A1.
PD 26-FEB-1998.
PF 20-AUG-1997; U14637.
PR 20-AUG-1996; US-700013.
PA (ALX ) ALLELIX NEUROSCIENCE INC.
PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;
DR WPI: 98-169158/15.
DR P-PSDB: W56362, W56363.
PT Human glycine transporter - useful for treating nervous system
PT disorders, e.g. pain, myoclonus, etc.
PS Example 1A; Pages 39-40; 90pp; English.
CC The present sequence encodes a fragment of the human Glyt-2 transporter.
CC It corresponds to nucleotides 39 to 1635 of the full length sequence
CC PCR (RT-PCR) from human spinal cord or cerebellum polyA mRNA, using
CC PCR primers V22939-40. The sequence differs slightly from the full length
CC Glyt-2 sequence because samples were collected from several donors.
CC Glyt-2 is a glycine transporter found predominantly in the brain stem
```

CC and the spinal cord. Glyt-2 regulates the levels of glycine, a major  
 CC neurotransmitter, that preferentially influences the activity of  
 CC N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine  
 CC receptors. The Glyt-2 protein can be used to treat or to diagnose a  
 CC nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm  
 CC or hyperactivity, epilepsy or spasticity associated with stroke, head  
 CC trauma, neuronal cell death, multiple sclerosis, spinal cord injury,  
 CC dystonia, Huntington's disease or amyotrophic lateral sclerosis.  
 SQ Sequence 1597 BP; 338 A; 457 C; 447 G; 355 T;

alignment\_scores:  
 Quality: 87.00 Length: 302  
 Ratio: 0.617 Gaps: 17  
 Percent Similarity: 46.689 Percent Identity: 22.517

alignment\_block:

us-09-060-609-2 x v22900 ..  
 gn seg 1/1 to: v22900 from: 1 to: 1597

```

25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla.....Le 39
|||||.....|
553 ACTGGACTTCATCTCCATGTTGGGTGGGTACGAGTGGGGTGGCAATG 602
|||||.....|
39 uLeuProPheSerLeuProLeuLeuGlyGlyGly.....51
|||||.....|
603 TCTGGAGGTTTCCCTACTGCGCTTCCAGACGGGGAGGTGCTTCTC 652
|||||.....|
52 .....SerGlySerGly.Glu.....56
|||||.....|
653 ATCCCTTACCTGATGCTGCTGCTGGGTGGATACCACTTCTTCTT 702
|||||.....|
57 .LysValSerValSerLysMetAlaAlaAlaTrpPro.....68
|||||.....|
703 GGAGGTCTCGTGGCGAGTTTGCAGCGGACCGAGTCTGTGTGA 752
|||||.....|
69 .....SerGlyProSerAlaProGluAlaValThrAla 79
|||||.....|
753 AGGCATCCAGCTCTACAGGCTGTGGCATCGGATGCTGATCAACTCT 802
|||||.....|
80 ArgLeuValGlyValLeuTrp.....86
|||||.....|
803 GTCTTAATAGCATATACATAAGTGTGATTATTGCTATACACTTTCTA 852
|||||.....|
87 .....PheValSerValThrGlyProTyrGlyAlaValAla 99
|||||.....|
853 CCGTCTTTCCTCTTGTGTCTGTA.....CCCTGGGCTCCTGCA 896
|||||.....|
99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp.....112
|||||.....|
897 ACAACCTTGGATACGCCAGAA.....TGCAAGATAAAACCAAA 937
|||||.....|
113 LeuLysValGlyGlnTyrIleCysLysAsp...ProLysIleAsnAspAl 128
|||||.....|
938 CTTTATTATGATCTCTGTATTATCAGTGACCATCCCAAAATA.....979
|||||.....|
128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysp 145
|||||.....|
980 .....CAGATCAAGACTGCTGCTTCTGCTGACCGCTT 1013
|||||.....|
145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
|||||.....|
1014 ATCCCAAGTGCACATGTTAATTCACCGAGGCGCAATACACATTT 1063
|||||.....|
162 PheThrGlyAsnGluValGlyPhePheLysProLysSerCysArgAsnVa 178
|||||.....|
1064 GTCAGTGAAGTGAG...GAGTACTTCAAGTACTTGTGCTGAAGATTTC 1110
|||||.....|
178 LAsnGlyTyrSerTyr.....LysValAlaValAlaLeuSerL 191
|||||.....|
1111 TGCAGGATTGAATATCTCGCGAGATCAGTGCCACTAGCTCTCTGCC 1160

```

```

191 euPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrProAla 207
|||||.....|
1161 TCTCTCTGGCTGGTGC.....ATTGTGATGATCG 1192
|||||.....|
208 Leu.....GlyLeuLeuLysPheCysThrValG 217
|||||.....|
1193 TTGGCTAAGGAATCAAGACTTCAGGAAAGTGGTACTTCCAGGCCAC 1242
|||||.....|
217 yPheCysGlyIleGlySerLeuIleAspPheIleLeuSerMetGlnI 234
|||||.....|
1243 GTTCCGCTATGTC...GTACTCGTATCTCTCATCCGAGGAGTCACCC 1289
|||||.....|
234 leValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrGlyThr 250
|||||.....|
1290 TGCCCTGGAGCTGGAGTCTGGTACTTCTCATCACACCCCAAGTGGGAG 1339
|||||.....|
251 ArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThrGlnLe 267
|||||.....|
1340 AAACCTACGGATGCCACGGTGTGGAAGATGCT.....GCCACTCAGAT 1383
|||||.....|
267 uTyr 268
|||||
1384 TTTC 1387

seq_name: N_Geneseq_36:T39750

seq_documentation_block:
ID T39750 standard; cDNA; 1624 BP.
AC T39750;
DT 01-JAN-1997 (first entry)
DE Human mucosal addressin cell adhesion molecule-1 cDNA clone 4.
KW Mucosal addressin cell adhesion molecule-1; MadCAM; antibody;
KW Leukocyte infiltration; ulcerative colitis; Crohn's disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..1221
FT signal_peptide 1..54 /*tag= a
FT FT /*tag= b
FT mat_peptide 55..1218
FT FT /*tag= c
PN W09624673-AL.
PD 15-AUG-1996.
PF 12-FEB-1996; U02153.
PR 10-FEB-1995; US-386857.
PR 01-SEP-1995; US-523004.
PA (LEUK-) LEUKOSITE INC.
PI Briskin MJ, Newman W, Picarella D, Ringler DJ;
DR WPI; 96-384445/38.
DR P-PSDB; W05321.
PT Nucleic acid encoding primate mucosal addressin cell adhesion
PT molecule-1 - and antibodies against this protein, useful to treat
PT diseases associated with leukocyte infiltration e.g. ulcerative
PT colitis
PS Claim 3; Page 104-106; 156pp; English.
CC cDNA clone 4 (T39750) codes for human mucosal addressin cell
CC adhesion molecule-1 (MadCAM-1) (W05321), whose cellular adhesion
CC function is alpha4beta7 integrin-dependent and pref. is selective
CC for alpha4beta7. It was isolated from a cDNA library prepd. from
CC normal human mesenteric lymph nodes (MLN) using a macaque MadCAM-1
CC cDNA clone (T39752) as probe. Another clone (T39751), encoding
CC a different human MadCAM isoform (W05322), was isolated from the
CC MLN of a patient with Crohn's disease. MadCAM nucleic acids can
CC be used as probes to detect polymorphic or allelic variants of
CC MadCAM, or used to produce recombinant MadCAMs in host cells.
SQ Sequence 1624 BP; 295 A; 610 C; 454 G; 265 T;

```

alignment\_scores:

Quality: 86.50 Length: 182  
 Ratio: 0.961 Gaps: 11  
 Percent Similarity: 49.451 Percent Identity: 24.725

```
alignment block:
US-09-060-609-2 x T39750/rev ..
Align seg 1/1 to reverse of: T39750 from: 1 to: 1624

      4 LeuLysGlySerProAsnValIleProArgAlaHis...GlyGlnLysAs 19
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1200 CTGGCCGGTCCCCCTTAACCCAGCCGCGACACCTCGGGGCGACAAGCC 1151

      19 nThrArgArgAspGlyThrGlyLeuTyrProMetArgGlyProPheLysA 36
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 TCAGAGAGCTGGTGGGTGGTCTGCTC..... 1123

      36 snLeuAlaLeuLeuProPheSerLeuPro...LeuLeuGlyGlyGly 51
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122 ..CTCAGCCAGGTGCCGCGCAGCTTTCCAGAGGTGATACGTGGCAAGGC 1075

      52 Ser.GlySerGlyLulysValSer.....ValSerLysMetAlaAlaA 66
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1074 CAGGAGCAGCAGTCCCGCAGCACCTGCTGGTCCACAGAGCCGCGGCA 1025

      66 laTrp...ProSerGlyProSerAlaProGluAlaValThrAlaArgLeu 81
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 GCTGTCACCCGCGAGTTGGACGAGCCTGTGGGATCACTTCTCCCTGC 975

      82 ValGlyValLeuTrpPheValSer.....Val.ThrThrG 93
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 GTGGGCCAGCCTGGGAGATCTCAGGGCGCGAGTCTGTGGAGCCTGG 925

      93 lyProTrpGlyAlaValAlaThrSerAlaGlyGlyGlu...GluSerLeu 108
      || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 GCTCCTGGGGGTGTGTGGAGCCCTGCTGGGGCGCGGCTCCGGGAGG 875

      109 LysCysGluAspLeuLysValGlyGlnTyrIle..... 119
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
874 TCTGTGGGAGCTCCG...GGGAGGTGGTGTGGGAGGCTCTCTGGAG 828

      120 .....CysLysAspProLysIleAsnA 127
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
827 GTGGTGTGGGAGGCTCCTGGGAGGTGGTGTGGGAGGCTCCTGGGAGGT 778

      127 spAlaThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSer 143
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 GGTGTGGGAGACTCCGGGAGGTGGTGTGGGAGACTCCGGGAGGTGG 728

      144 CysPheProAlaProAsnIleThrCysLysAspSerSerGly 157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 TGTGGGAGCTCCGGGAGGTGGTGTGGGAGGCTCCCGGG 686
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 13:23:38 : Search time 42.25 Seconds  
(without alignments)  
2220.775 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTAAAGGTC.....AAACGCAATATATCCATAA 810

Scoring table: IDENTITY\_NUC

Searched: 207703 seqs, 57918364 residues

Database: Issued\_Patents\_NA:\*

d size: 0

Number of hits that pass the threshold : 415406  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS9\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.4	4.5	1462	2	US-08-552-142A-16
2	36.4	4.5	1951	2	US-08-910-973-16
3	34.6	4.3	4673	1	US-07-638-431-1
4	34.6	4.3	4673	5	PCT-US92-00018-1
5	33.4	4.1	1636	6	5447867-2
6	32.8	4.0	988	1	US-08-243-545-5
7	32.8	4.0	1700	2	US-08-539-304A-3
8	32.8	4.0	988	3	US-08-993-962-5
9	32.8	4.0	988	5	PCT-US94-05365-5
10	32.4	4.0	31571	1	US-08-323-443B-1
11	32.2	4.0	3012	3	US-08-475-427-5
12	32.2	4.0	3012	4	US-07-842-165-5
13	31.6	3.9	532	1	US-08-558-735-1
14	31.6	3.9	345	1	US-08-558-735-3
15	31.6	3.9	345	1	US-08-558-735-16
16	31.6	3.9	345	1	US-08-558-735-17
17	31.6	3.9	333	1	US-08-558-735-18
18	31.6	3.9	330	1	US-08-558-735-19
19	31.6	3.9	857	5	PCT-US94-05150-19
20	31.4	3.9	8700	3	US-08-392-625-16
21	31.4	3.9	8700	3	US-08-466-961A-16
22	31.4	3.9	8700	4	US-08-645-193B-18
23	31.4	3.9	8298	5	PCT-US93-03076-1
24	31.2	3.9	1644	3	US-08-458-553-1
25	31	3.8	2688	4	US-08-909-965C-1
26	30.6	3.8	440	1	US-08-558-735-4
27	30.6	3.8	345	1	US-08-558-735-6
28	30.6	3.8	522	1	US-08-558-735-7
29	30.6	3.8	342	1	US-08-558-735-9
30	30.4	3.8	3933	1	US-08-199-776-1
31	30.4	3.8	901	2	US-08-508-448C-9
32	30.4	3.8	3933	5	PCT-US95-02044-1
33	30.2	3.7	3143	2	US-08-485-621-1

C 34	30.2	3.7	3143	4	US-08-973-831-1	Sequence 1, Appli
C 35	30.2	3.7	3143	5	PCT-US96-09530A-1	Sequence 1, Appli
C 36	30	3.7	8438	1	US-07-945-283-1	Sequence 1, Appli
C 37	29.8	3.7	996	4	US-08-924-759-21	Sequence 21, Appli
C 38	29.6	3.7	1186	4	US-08-731-722-5	Sequence 5, Appli
C 39	29.4	3.6	1288	2	US-08-440-856A-9	Sequence 9, Appli
C 40	29.4	3.6	1517	2	US-08-508-448C-15	Sequence 15, Appli
C 41	29.2	3.6	3480	1	US-07-657-769B-68	Sequence 68, Appli
C 42	29.2	3.6	17041	1	US-08-076-011-1	Sequence 1, Appli
C 43	29.2	3.6	30001	1	US-08-125-468-1	Sequence 1, Appli
C 44	29.2	3.6	3480	1	US-07-789-184-219	Sequence 219, App
C 45	29.2	3.6	3480	2	US-08-475-263-219	Sequence 219, App

RESULT 1  
US-08-552-142A-16  
Sequence 16, Application US/08552142A  
Patent No. 569595  
GENERAL INFORMATION:  
APPLICANT: Weintraub, Harold M.  
APPLICANT: Lee, Jacqueline E.  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Hollenberg, Stanley M.  
TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes  
TITLE OF INVENTION: and Proteins  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson KindnessPLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/552,142A  
FILING DATE: 02-NOV-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FHCR-1-8933  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-225-0709  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1462 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE: Mus musculus  
IMMEDIATE SOURCE:  
CLONE: i.1.1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 231..1101  
US-08-552-142A-16

ALIGNMENTS



```
;; ORIGINAL SOURCE:
;; ORGANISM: Plasmodium yoelii
;; STRAIN: 17X(NL)
;; DEVELOPMENTAL STAGE: erythrocytic stage
;; TISSUE TYPE: Blood
;; CELL TYPE: erythrocytic stage
;; IMMEDIATE SOURCE:
;; LIBRARY: Py-lambdaagtl1-2-7 kb genomic expression
;; CLONE: Py10.1111
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 718..3195
;; OTHER INFORMATION:
US-07-638-431-1

Query Match          4.3%; Score 34.6; DB 1; Length 4673;
Best Local Similarity 55.4%; Pred. No. 0.34; Mismatches 0; Gaps 0;
Matches 67; Conservative 0; Indels 54; Indels 0; Gaps 0;

Db 560 TCGCATGTCTCTTTTCTTGGATGTTGGAGCAGATCGATTTTACCTTGGATACCCCTG 619
1993 TTGGCTTGTGTTGGGTTGATGTTGCTTTGGGTTGGCTTGTGCTTGGCTTGTG 1934

Qy 620 CTTTGGGTTGTTAAAGTTTTCACGTAGGTTTGTGGAATTTGGAGCCCTAATTGATT 679
1933 GTTTGGCTTGTGTTGGGTTTCGCCCTTTTGGGTTTCGCCCTTTTGGGTTGGT 1874

Qy 680 T 680
Db 1873 T 1873

RESULT 4
PCT-US92-00018-1/c
; Sequence 1, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: N
;; ANTI-SENSE: N
;; ORIGINAL SOURCE:
;; ORGANISM: Plasmodium yoelii
;; STRAIN: 17X(NL)
;; DEVELOPMENTAL STAGE: erythrocytic stage
;; TISSUE TYPE: Blood
;; CELL TYPE: erythrocytic stage
;; IMMEDIATE SOURCE:
;; LIBRARY: Py-lambdaagtl1-2-7 kb genomic expression
;; CLONE: Py10.1111
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 718..3195
;; OTHER INFORMATION:
PCT-US92-00018-1

Query Match          4.3%; Score 34.6; DB 5; Length 4673;
Best Local Similarity 55.4%; Pred. No. 0.34; Mismatches 0; Gaps 0;
Matches 67; Conservative 0; Indels 54; Indels 0; Gaps 0;

Qy 560 TCGCATGTCTCTTTTCTTGGATGTTGGAGCAGATCGATTTTACCTTGGATACCCCTG 619
1993 TTGGCTTGTGTTGGGTTGATGTTGCTTTGGGTTGGCTTGTGCTTGGCTTGTG 1934

Qy 620 CTTTGGGTTGTTAAAGTTTTCACGTAGGTTTGTGGAATTTGGAGCCCTAATTGATT 679
1933 GTTTGGCTTGTGTTGGGTTTCGCCCTTTTGGGTTTCGCCCTTTTGGGTTGGT 1874

Qy 680 T 680
Db 1873 T 1873

RESULT 5
5447867-2
; Patent No. 5447867
; APPLICANT: BRIDGES, TAN;SCHUCH, WOLFGANG;GRIERSON, DONALD
; TITLE OF INVENTION: RECOMBINANT DNA CONTAINING PECTIN
; ESTERASE GENE SEGMENTS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/24,866
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 720,629
; FILING DATE: 25-JUN-1991
; APPLICATION NUMBER: 419,779
; FILING DATE: 29-SEP-1989
; APPLICATION NUMBER: 119,614
; FILING DATE: 12-NOV-1987
; SEQ ID NO:2:
; LENGTH: 1636
5447867-2

Query Match          4.1%; Score 33.4; DB 6; Length 1636;
Best Local Similarity 58.6%; Pred. No. 0.43; Mismatches 0; Gaps 0;
Matches 58; Conservative 0; Indels 41; Indels 0; Gaps 0;

Qy 687 TATTCAATGCAGATTTTGGACCTTCAGATGGAAGTAGTACATTATAGATTACTATGG 746
Db 449 tatttcagtaagatttttggatccttagaagcatcagtaaaatttcagactacaaga 508

Qy 747 AACGACATTCACAGACTGAGTATTACTAATGAACATT 785
Db 509 tagaagccttggtgattgcttttgatagtagttcaaaattt 547

RESULT 6
US-08-243-545-5/c
```

RESULT 7  
 US-08-539-304A-3  
 ; Sequence 3, Application US/08539304A  
 ; Patent No. 5792933  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MA, DIN-POW  
 ; TITLE OF INVENTION: FIBER-SPECIFIC PROTEIN EXPRESSION IN THE  
 ; TITLE OF INVENTION: COTTON PLANT  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT  
 ; STREET: 1755 JEFFERSON DAVIS HWY. SUITE 400  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/539.304A  
 ; FILING DATE: 04-OCT-1995  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: NORMAN, OBLON F  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 2343-037-27  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1700 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 801..936  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: join(448..800, 937..943)  
 ; US-08-539-304A-3

```

Query Match      4.0%; Score 32.8; DB 2; Length 1700;
Best Local Similarity. 58.0%; Pred. NO. 0.69;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY   631  TTAAGTTTTGCACCTAGGTTTTGTGGAAATGGGAGCCATAATTGATTCATTCTATT 690
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   1562 TTATGATTTATGATGATAATATTATAGCATTTATTAGCATAAATATATTGAATTGATT 1621
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   691  TCAATGCAGATTTGTGACCCTTCAGATGGAAAGTAGTTACA 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   1622 AGAATTAGGTTTTTTCACCTATAATAGATGTAGTCAA 1661
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT      8
US-08-993-962-5/C
; Sequence 5, Application US/08993962
; Patent No. 5843423
; GENERAL INFORMATION:
; APPLICANT: Lyman, Stewart D.
; APPLICANT: Beckmann, M. Patricia
; TITLE OF INVENTION: Ligands for flt3/flk-2 Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Stephen L. Muaska, Immunex Corporation
; STREET: 51 University Street

```

	Query Match	4.0%	Score 32.8;	DB 1;	Length 988;
	Best Local Similarity	56.5%;	Pred. No. 0.49;		
	Matches 61;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;
QY	127	TCCTCCCGTCTCTGGCGGAGCGCGGAAGCGGAGTGGCGAGAAAGTGTCCGTTCCAA	186		
Db	108	TCCCACTGAGTCCCGAGCTCAGCAGCAGCAGCAGGAGGAGATAGTTGTTGGGCTCCAG	49		
QY	187	ATGGCGGCGCCTTGGCGGTCTGGTCCGCTGCTCCGGAGGCCCGTGAC	234		
Db	48	CTGGCGCCAGCACTGTCAATTTCGGCGGGGCGCCCGGAATTCGGCGCC	1		

CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.0.1  
SOFTWARE: Microsoft Word, Version #5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,962  
FILING DATE: December 18, 1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/162,407  
FILING DATE: December 3, 1993  
APPLICATION NUMBER: 08/111,758  
FILING DATE: August 25, 1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/068,394  
FILING DATE: May 24, 1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Malaska, Stephen L.  
REGISTRATION NUMBER: 32,655  
REFERENCE/DOCKET NUMBER: 2813-C  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 988 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..734  
PCT-US94-05365-5

Query Match 4.0%; Score 32.8; DB 3; Length 988;  
Best Local Similarity 56.5%; Pred. No. 0.49; Mismatches 0; Gaps 0;  
Matches 61; Conservative 0; Indels 47; Indels 0; Gaps 0;  
QY 127 TCCCTCCCGCTCTGGCGGAGCGGGAAGTGGCGAGAAAGTGTGGTCTCCAAAG 186  
Db 108 TCCCACTAGTCCCGAGTCCAGCAGCAGCAGGAGATAGTGTGGCTCCAGG 49  
QY 187 ATGGCGCGCGCTGGCGTGTGGTCTGTCTCCGAGGCCGTGACG 234  
Db 48 CTGGCGCCAGCACTGTCTATTTCCGCGCGCGCGCGGAAATTCGGCGCG 1

RESULT 9  
PCT-US94-05365-5/c  
Sequence 5, Application PC/TUS9405365  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
TITLE OF INVENTION: Ligands for flt3/flk-2 Receptors  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stephen L. Malaska, Immunex Corporation

STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05365  
FILING DATE: May 24, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: -to be assigned-  
FILING DATE: May 11, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/209,502  
FILING DATE: March 7, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/162,407  
FILING DATE: December 3, 1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/111,758  
FILING DATE: August 25, 1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/106,463  
FILING DATE: August 12, 1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/068,394  
FILING DATE: May 24, 1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Malaska, Stephen L.  
REGISTRATION NUMBER: 32,655  
REFERENCE/DOCKET NUMBER: 2813-B  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 988 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..734  
PCT-US94-05365-5

Query Match 4.0%; Score 32.8; DB 5; Length 988;  
Best Local Similarity 56.5%; Pred. No. 0.49; Mismatches 0; Gaps 0;  
Matches 61; Conservative 0; Indels 47; Indels 0; Gaps 0;  
QY 127 TCCCTCCCGCTCTGGCGGAGCGGGAAGTGGCGAGAAAGTGTGGTCTCCAAAG 186  
Db 108 TCCCACTAGTCCCGAGTCCAGCAGCAGCAGGAGATAGTGTGGCTCCAGG 49  
QY 187 ATGGCGCGCGCTGGCGTGTGGTCTGTCTCCGAGGCCGTGACG 234  
Db 48 CTGGCGCCAGCACTGTCTATTTCCGCGCGCGCGCGGAAATTCGGCGCG 1



```

: NAME/KEY: mat_peptide
: LOCATION: join(2157..2383, 2463..2616, 2698..3007, 2698)
: US-07-842-165-5

Query Match 4.0%; Score 32.2; DB 4; Length 3012;
Best Local Similarity 63.6%; Pred. No. 1.6;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 717 TGGAAAGTAGTACATTATAGATTACTATGGAACAGACTTACAGACTGAGTATTACTAA 776
Db 920 TTGAATTTTGAATTATGAAATTATATGATTCAGACTTCAAGAGATTCTTTAATACGGA 861
QY 777 TGAACACATTAGAAAAA 793
Db 860 TAAAGAATTTTGACAAA 844

RESULT 13
US-08-558-735-1/c
: Sequence 1, Application US/08558735
: Patent No. 5681724
: GENERAL INFORMATION:
: APPLICANT: Tripp, Cynthia A.
: APPLICANT: Brandt, Kevin S.
: APPLICANT: Wisniewski, Nancy
: TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION
: TITLE OF INVENTION: INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, A
: NUMBER OF SEQUENCES: 23
: NUMBER OF INVENTION: USES THEREOF
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/558,735
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-43
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 532 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 8..355
: US-08-558-735-1

Query Match 3.9%; Score 31.6; DB 1; Length 532;
Best Local Similarity 54.2%; Pred. No. 0.8;
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 390 AGAACCAAGTTAACTGTACAAACTACACAGCTCATGTTTCTCTGTTTCCAGCACCCACAT 449
Db 208 AGAACCAAGTTGATTTTAAACACACACAGCATGAGTCTGTTCTACTTCCACGGAATGTCAT 149

```

QY 450 AACTTGAAGATTCCAGTGCATGAACACACATTTTACTGGGACGAAAGTGGTTTT 507  
Db 148 CGCTTGTCACCATTCACATGGATTGATACGTAATTTCCGGTTTTTCCAAAGTGGTTTT 91

RESULT 14  
US-08-558-735-3/c  
; Sequence 3, Application US/08558735  
; Patent No. 5681724  
; GENERAL INFORMATION:  
; APPLICANT: Tripp, Cynthia A.  
; APPLICANT: Brandt, Kevin S.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION  
; TITLE OF INVENTION: INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/558,735  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-43  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-558-735-3

Query Match 3.9%; Score 31.6; DB 1; Length 345;  
Best Local Similarity 54.2%; Pred. No. 0.61;  
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 390 AGAACCGATTAACTTAACTACAACTACACAGCTCATGTTTCTCTGTTTCCAGCACCCACAT 449  
Db 201 AGAACCGATTGATTTTAAACACACACAGCAGCATGGATCTGTACTTCCACCGAATGTCAT 142

QY 450 AACTTGAAGATTCCAGTGCATGAACACACATTTTACTGGGACGAAAGTGGTTTT 507  
Db 141 CGCTTGTCACCATTCACATGGATTGATACGTAATTTCCGGTTTTTCCAAAGTGGTTTT 84

RESULT 15  
US-08-558-735-16/c  
; Sequence 16, Application US/08558735  
; Patent No. 5681724  
; GENERAL INFORMATION:  
; APPLICANT: Tripp, Cynthia A.  
; APPLICANT: Brandt, Kevin S.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: PARASITIC HELMINTH MACROPHAGE MIGRATION

; TITLE OF INVENTION: INHIBITORY FACTOR PROTEINS, NUCLEIC ACID MOLECULES, AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/558,735  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-43  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 355 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-558-735-16

Query Match 3.9%; Score 31.6; DB 1; Length 355;  
Best Local Similarity 54.2%; Pred. No. 0.62;  
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 390 AGAACCGATTAACTTAACTACAACTACACAGCTCATGTTTCTCTGTTTCCAGCACCCACAT 449  
Db 208 AGAACCGATTGATTTTAAACACACACAGCAGCATGGATCTGTACTTCCACCGAATGTCAT 149

QY 450 AACTTGAAGATTCCAGTGCATGAACACACATTTTACTGGGACGAAAGTGGTTTT 507  
Db 148 CGCTTGTCACCATTCACATGGATTGATACGTAATTTCCGGTTTTTCCAAAGTGGTTTT 91

Search completed: February 10, 2000, 13:25:30  
Job time: 112 sec





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2000, 17:55:13 ; Search time 23.51 Seconds  
(without alignments)  
147.013 Million cell updates/sec

Title: US-09-060-609-2

Perfect score: 269

Sequence: 1 MHILKSPNVPRAHGOKT.....TTLRLSITNETFRKTQLYP 269

Scoring table: OLIGO

Searched: 130275 seqs, 12848600 residues

Database: Issued\_Patents\_AA:\*

Database size: 0

Number of hits that pass the threshold : 130275  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	2.6	452	1	US-08-290-978A-5
2	7	2.6	302	1	US-08-121-054C-18
3	7	2.6	302	1	US-08-121-054C-30
4	7	2.6	211	1	US-08-276-852-34
5	7	2.6	211	1	US-08-133-011-16
6	7	2.6	482	1	US-08-358-160-5
7	7	2.6	484	1	US-08-358-160-7
8	7	2.6	399	1	US-08-530-950-10
9	7	2.6	211	1	US-08-322-730A-16
10	7	2.6	211	1	US-08-387-874-16
11	7	2.6	211	1	US-08-899-575-34
12	7	2.6	211	1	US-08-899-575-34
13	7	2.6	1940	2	US-08-644-271-30
14	7	2.6	461	2	US-08-463-587A-26
15	7	2.6	452	2	US-08-780-869-5
16	7	2.6	461	2	US-08-463-667A-4
17	7	2.6	246	2	US-08-553-497A-24
18	7	2.6	1186	2	US-08-861-464-8
19	7	2.6	1186	2	US-08-396-001-8
20	7	2.6	745	2	US-08-674-887A-8
21	7	2.6	211	2	US-08-383-619-16
22	7	2.6	233	2	US-08-829-876-101
23	7	2.6	232	2	US-08-829-876-103
24	7	2.6	233	2	US-08-829-876-105
25	7	2.6	534	2	US-08-691-814B-8
26	7	2.6	399	2	US-08-874-186-92
27	7	2.6	461	3	PCT-US91-09133-27
28	7	2.6	211	3	PCT-US93-08364-16
29	7	2.6	293	3	PCT-US94-05669A-4
30	7	2.6	293	3	PCT-US94-05669A-6
31	7	2.6	349	3	PCT-US95-01185-139
32	7	2.6	301	3	PCT-US95-01185-142
33	7	2.6	335	3	PCT-US95-01185-143
34	7	2.6	337	3	PCT-US95-01185-148
35	7	2.6	349	3	PCT-US95-01185-151

ALIGNMENTS

RESULT 1

US-08-290-978A-5  
; Sequence 5, Application US/08290978A  
; Patent No. 5624834

GENERAL INFORMATION:

APPLICANT: KUSTERS-VAN SOMEREN, MARGO A.

APPLICANT: MULLER, YVONNE

APPLICANT: KESTER, HERMANUS C.M.

APPLICANT: VISSER, JACOB

APPLICANT: VAN OYEN, ALBERT J.J.

APPLICANT: ROLIN, CLAUDE

TITLE OF INVENTION: CLONING AND EXPRESSION OF THE

TITLE OF INVENTION: EXO-POLYGALACTURONASE GENE FROM ASPERGILLUS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 Pennsylvania Avenue N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1812

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,978A

FILING DATE: 17-OCT-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 4615-0044.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763

TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 452 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-290-978A-5

Query Match 2.6%; Score 7; DB 1; Length 452;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGS 54

DB 336 GGGGSGS 342

RESULT 2

US-08-121-054C-18

; Sequence 18, Application US/08121054C  
; Patent No. 5637481  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Gilliland, Lisa K.  
; APPLICANT: Hayden, Martha S.  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Bajorath, Jurgen  
; APPLICANT: Fell, Perry  
; TITLE OF INVENTION: Expression Vectors Encoding Bispecific  
; TITLE OF INVENTION: Fusion Proteins and Methods of Producing Biologically  
; TITLE OF INVENTION: Active Bispecific Fusion Proteins in a Mammalian Cell  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,054C  
; FILING DATE: 13-SEP-1993  
; CLASSIFICATION: 435  
; APPLICATION NUMBER: US 08/013,420  
; FILING DATE: 01-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436.18US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 302 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: linear  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-121-054C-18

Query Match 2.6%; Score 7; DB 1; Length 302;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 LGGGGSG 53  
Db 133 LGGGGSG 139

RESULT 3  
US-08-121-054C-30  
; Sequence 30, Application US/08121054C  
; Patent No. 5637481  
; GENERAL INFORMATION:  
; APPLICANT: Ledbetter, Jeffrey A.  
; APPLICANT: Gilliland, Lisa K.  
; APPLICANT: Hayden, Martha S.  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Bajorath, Jurgen  
; APPLICANT: Fell, Perry  
; TITLE OF INVENTION: Expression Vectors Encoding Bispecific  
; TITLE OF INVENTION: Fusion Proteins and Methods of Producing Biologically  
; TITLE OF INVENTION: Active Bispecific Fusion Proteins in a Mammalian Cell  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant & Gould  
; STREET: 11150 Santa Monica Blvd., Suite 400  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,054C  
; FILING DATE: 13-SEP-1993  
; CLASSIFICATION: 435  
; APPLICATION NUMBER: US 08/013,420  
; FILING DATE: 01-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adriano, Sarah B.  
; REGISTRATION NUMBER: 34,470  
; REFERENCE/DOCKET NUMBER: 30436.18US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 310-445-1140  
; TELEFAX: 310-445-9031  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-121-054C-30

Query Match 2.6%; Score 7; DB 1; Length 302;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 LGGGGSG 53  
Db 133 LGGGGSG 139

RESULT 4  
US-08-276-852-34  
; Sequence 34, Application US/08276852  
; Patent No. 5652138  
; GENERAL INFORMATION:  
; APPLICANT: Burton, Dennis R  
; APPLICANT: Barbas, Carlos F  
; APPLICANT: Lerner, Richard A  
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS  
; NUMBER OF SEQUENCES: 170  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,  
; STREET: Mail Drop TPC8  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,852  
; FILING DATE: 18-JUL-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,302  
FILING DATE: 30-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/954,148  
FILING DATE: 30-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: SCRL452P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 211 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
US-08-133-011-16  
8-276-852-34

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGSG 55  
Db 55 GGGSGSG 61  
|||||||

RESULT 5  
US-08-133-011-16  
Sequence 16, Application US/08133011  
Patent No. 5658727  
GENERAL INFORMATION:  
APPLICANT: Kang, Angray  
APPLICANT: Barbas, Carlos  
APPLICANT: Lerner, Richard A.  
TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING  
TITLE OF INVENTION: PHAGEMIDS  
NUMBER OF SEQUENCES: 161  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5638727th Torrey Pines Road, TPC-8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/133,011  
FILING DATE: 08-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/683,602  
FILING DATE: 10-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/826,623  
FILING DATE: 27-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US 92/03091  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: SCRF 238.2

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 211 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
US-08-133-011-16

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGSG 55  
Db 55 GGGSGSG 61  
|||||||

RESULT 6  
US-08-358-160-5  
Sequence 5, Application US/08358160  
Patent No. 5663143  
GENERAL INFORMATION:  
APPLICANT: LEY, Arthur C.  
APPLICANT: LADNER, Robert C.  
APPLICANT: GUTERMAN, Sonia K.  
APPLICANT: ROBERTS, Bruce L.  
APPLICANT: MARKLAND, William  
APPLICANT: KENT, Rachel B.  
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE  
NUMBER OF SEQUENCES: 234  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W. Suite 300  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,160  
FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319  
FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/664,989  
FILING DATE: 01-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,063  
FILING DATE: 02-MAR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/240,160  
FILING DATE: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 482 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-5

Query Match 2.6%; Score 7; DB 1; Length 482;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

49 GGGSGG 55  
|||||||  
326 GGGSGG 332

RESULT 7  
US-08-358-160-7  
Sequence 7, Application US/08358160  
Patent No. 5663143  
GENERAL INFORMATION:  
APPLICANT: LEY, Arthur C.  
APPLICANT: LADNER, Robert C.  
APPLICANT: GUTERMAN, Sonia K.  
APPLICANT: ROBERTS, Bruce L.  
APPLICANT: MARKLAND, William  
APPLICANT: KENT, Rachel B.  
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
NUMBER OF SEQUENCES: 234  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W. Suite 300  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,160  
FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319  
FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/664,989  
FILING DATE: 01-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,063  
FILING DATE: 02-MAR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/240,160  
FILING DATE: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 484 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-7

Query Match 2.6%; Score 7; DB 1; Length 484;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGG 55  
|||||||  
Db 328 GGGSGG 334

RESULT 8  
US-08-530-950-10  
Sequence 10, Application US/08530950  
Patent No. 5736381  
GENERAL INFORMATION:  
APPLICANT: Davis, Roger J.  
APPLICANT: Raingeaud, Joel  
APPLICANT: Gupta, Shashi  
APPLICANT: Derijard, Benoit  
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND  
ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE  
TITLE OF INVENTION: KINASES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,950  
FILING DATE: 19-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07917/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 399 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
US-08-530-950-10

Query Match 2.6%; Score 7; DB 1; Length 399;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGG 55  
|||||||

Db 13 GGGSGG 19

RESULT 9  
US-08-322-730A-16  
; Sequence 16, Application US/08322730A  
; Patent No. 5759817  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos  
; TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; NUMBER OF SEQUENCES: 125  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10550 No. 5759817th Torrey Pines Road, Suite 220,  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08322,730A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/826,623  
; FILING DATE: 27-JAN-1992  
; APPLICATION NUMBER: US 07/683,602  
; FILING DATE: 10-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bingham, Douglas A  
; REGISTRATION NUMBER: 32,457  
; REFERENCE/DOCKET NUMBER: SCR0707P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-784-2937  
; TELEFAX: 619-784-9399

INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 211 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
US-08-322-730A-16

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGG 55  
Db 55 GGGSGG 61

RESULT 10  
US-08-387-874-16  
; Sequence 16, Application US/08387874  
; Patent No. 5770356  
; GENERAL INFORMATION:  
; APPLICANT: Light, Paul L., II  
; APPLICANT: Lerner, Richard A.  
; TITLE OF INVENTION: PHAGMIDS EXPRESSING A SURFACE  
; TITLE OF INVENTION: RECEPTOR AND A SURFACE HETEROLOGOUS PROTEIN  
; NUMBER OF SEQUENCES: 97  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5770356th Torrey Pines Road, Suite 220,  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08387,874  
; FILING DATE: 22-FEB-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/08364  
; FILING DATE: 03-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/941,369  
; FILING DATE: 04-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSRI 303.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-554-2937  
; TELEFAX: 619-554-6312  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 211 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
US-08-387-874-16

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGG 55  
Db 55 GGGSGG 61

RESULT 11  
US-08-899-575-34  
; Sequence 34, Application US/08899575  
; Patent No. 5770440  
; GENERAL INFORMATION:  
; APPLICANT: Burton, Dennis R  
; APPLICANT: Barbas, Carlos F  
; APPLICANT: Lerner, Richard A  
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS  
; NUMBER OF SEQUENCES: 170  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/899,575  
;; FILING DATE: 24-JUL-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/276,852  
;; FILING DATE: 18-JUL-1994  
;; APPLICATION NUMBER: US 08/178,302  
;; FILING DATE: 30-SEP-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/954,148  
;; FILING DATE: 30-SEP-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fitting, Thomas  
;; REGISTRATION NUMBER: 34,163  
;; REFERENCE/DOCKET NUMBER: SCRI452P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-554-2937  
;; TELEFAX: 619-554-6312  
;; INFORMATION FOR SEQ ID NO: 34:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 211 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; FRAGMENT TYPE: internal  
;; US-08-899-575-34

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 49 GGGSGSG 55  
Db 55 GGGSGSG 61

RESULT 12  
US-08-899-575-34  
; Sequence 34, Application US/08899575  
; Patent No. 5804440  
; GENERAL INFORMATION:  
; APPLICANT: Burton, Dennis R  
; APPLICANT: Barbas, Carlos F  
; APPLICANT: Lerner, Richard A  
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
; NUMBER OF SEQUENCES: 170  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10666 No. 5804440th Torrey Pines Road, Suite 220,  
; STREET: Mail Drop TPC8  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/899,575  
; FILING DATE: 24-JUL-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/276,852  
; FILING DATE: 18-JUL-1994  
; APPLICATION NUMBER: US 08/178,302

;; FILING DATE: 30-SEP-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/954,148  
;; FILING DATE: 30-SEP-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fitting, Thomas  
;; REGISTRATION NUMBER: 34,163  
;; REFERENCE/DOCKET NUMBER: SCRI452P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-554-2937  
;; TELEFAX: 619-554-6312  
;; INFORMATION FOR SEQ ID NO: 34:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 211 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; FRAGMENT TYPE: internal  
;; US-08-899-575-34

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 49 GGGSGSG 55  
Db 55 GGGSGSG 61

RESULT 13  
US-08-644-271-30  
; Sequence 30, Application US/08644271  
; Patent No. 5814478  
; GENERAL INFORMATION:  
; APPLICANT: Valenzuela, et al.  
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/644,271  
; FILING DATE: 10-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 60/008,657  
; FILING DATE: 15-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cobert, Robert J  
; REGISTRATION NUMBER: 36,108  
; REFERENCE/DOCKET NUMBER: REG 195A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-345-7400  
; TELEFAX: 914-345-7721  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1940 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Rat Agrin  
LOCATION: 1...1940  
OTHER INFORMATION:  
US-08-644-271-30

Query Match 2.6%; Score 7; DB 2; Length 1940;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GSGSGS 56  
| | | | |  
DB 567 GSGSGS 573

## RESULT 14

US-08-463-587A-26  
Sequence 26, Application US/08463587A  
Patent No. 5821047

## GENERAL INFORMATION:

APPLICANT: Garrard, Lisa J.  
APPLICANT: Henner, Dennis J.  
APPLICANT: Bass, Steven  
APPLICANT: Greene, Ronald  
APPLICANT: Lowman, Henry B.  
APPLICANT: Wells, James A.  
APPLICANT: Matthews, David J.

TITLE OF INVENTION: ENRICHMENT METHOD FOR VARIANT PROTEINS WITH  
TITLE OF INVENTION: ALTERED BINDING PROPERTIES

NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,587A  
FILING DATE: 05-JUN-1995

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/050058  
FILING DATE: 30-APR-1993  
APPLICATION NUMBER: PCT/US91/09133

FILING DATE: 03-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/743614

FILING DATE: 09-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/715300

FILING DATE: 14-JUN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/583400

FILING DATE: 10-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/621667

FILING DATE: 03-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: Schwartz, Timothy R.

REGISTRATION NUMBER: 32171

REFERENCE/DOCKET NUMBER: P0645P4D2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-7467

TELEFAX: 415/952-9881

TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 461 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-463-587A-26

Query Match 2.6%; Score 7; DB 2; Length 461;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GSGSGS 55  
| | | | |  
DB 305 GSGSGS 311

## RESULT 15

US-08-780-869-5

Sequence 5, Application US/08780869

Patent No. 5830737

## GENERAL INFORMATION:

APPLICANT: KUSTERS-VAN SOMEREN, MARGO A.  
APPLICANT: MULLER, YVONNE  
APPLICANT: KESTER, HERMANUS C.M.  
APPLICANT: VISSER, JACOB  
APPLICANT: VAN Ooyen, ALBERT J.J.  
APPLICANT: ROLIN, CLAUDE

TITLE OF INVENTION: CLONING AND EXPRESSION OF THE

TITLE OF INVENTION: EXO-POLYGALACTURONASE GENE FROM ASPERGILLUS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 Pennsylvania Avenue N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1812

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/780,869  
FILING DATE: 24-JAN-1997

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/290,978

FILING DATE: 17-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 4615-0044.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763

TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 452 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-780-869-5

Query Match 2.6%; Score 7; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GSGSGS 54



Db 336 GGGSGS 342  
|||||||

Search completed: February 10, 2000, 18:03:22  
Job time: 489 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 22:00:30 ; Search time 42.58 Seconds  
(without alignments)  
2203.564 Million cell updates/sec

Title: US-09-060-609-1

Perfect score: 810

Sequence: 1 ATGCATATTTTAAAGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: OLIGO\_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued Patents\_NA:\*

Size : 0

Number of hits that pass the threshold : 415406

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS9\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	17	2.1	3383	1 US-07-707-367-1	Sequence 1, Appli
2	17	2.1	8438	1 US-07-945-283-1	Sequence 1, Appli
3	17	2.1	3306	1 US-08-261-206A-71	Sequence 71, Appl
4	17	2.1	1251	1 US-08-261-206A-75	Sequence 75, Appl
5	17	2.1	2625	3 US-08-759-945-1	Sequence 1, Appli
6	17	2.1	1299	3 US-08-994-719C-3	Sequence 3, Appli
7	16	2.0	1393	1 US-08-052-205-2	Sequence 2, Appli
8	16	2.0	1451	1 US-08-052-205-3	Sequence 3, Appli
9	16	2.0	1451	1 US-08-031-143B-68	Sequence 68, Appl
10	16	2.0	301	1 US-08-031-143B-76	Sequence 76, Appl
11	16	2.0	2542	1 US-08-120-960-1	Sequence 1, Appli
12	16	2.0	966	1 US-08-162-475A-1	Sequence 1, Appli
13	16	2.0	935	1 US-08-162-475A-3	Sequence 3, Appli
14	16	2.0	1393	2 US-08-595-974-2	Sequence 2, Appli
15	16	2.0	1470	2 US-08-595-974-3	Sequence 3, Appli
16	16	2.0	6453	2 US-08-306-691B-14	Sequence 14, Appl
17	16	2.0	26	2 US-08-244-488-3	Sequence 3, Appli
18	16	2.0	1473	2 US-08-672-571A-2	Sequence 2, Appli
19	16	2.0	1386	2 US-08-672-571A-4	Sequence 4, Appli
20	16	2.0	10898	4 US-08-481-658B-5	Sequence 5, Appli
21	16	2.0	3680	4 US-08-494-907-1	Sequence 1, Appli
22	16	2.0	5076	4 US-08-494-907-2	Sequence 2, Appli
23	16	2.0	6387	4 US-08-494-907-3	Sequence 3, Appli
24	16	2.0	6170	4 US-08-494-907-4	Sequence 4, Appli
25	16	2.0	1047	4 US-08-494-907-11	Sequence 11, Appl
26	16	2.0	10898	4 US-08-477-504A-5	Sequence 5, Appli
27	16	2.0	10898	4 US-08-486-756A-5	Sequence 5, Appli
28	16	2.0	10898	4 US-08-485-862B-5	Sequence 5, Appli
29	16	2.0	3680	5 PCT-US96-10986-1	Sequence 1, Appli
30	16	2.0	5076	5 PCT-US96-10986-2	Sequence 2, Appli
31	16	2.0	6387	5 PCT-US96-10986-3	Sequence 3, Appli
32	16	2.0	6170	5 PCT-US96-10986-4	Sequence 4, Appli
33	16	2.0	1047	5 PCT-US96-10986-11	Sequence 11, Appli

C 34	16	2.0	1730	6 5223391-8	Patent No. 5223391
C 35	15	1.9	2516	1 US-07-914-282D-2	Sequence 2, Appli
C 36	15	1.9	4380	1 US-07-582-945-1	Sequence 1, Appli
C 37	15	1.9	2608	1 US-07-725-083-1	Sequence 1, Appli
C 38	15	1.9	2051	1 US-07-864-475A-3	Sequence 3, Appli
C 39	15	1.9	2516	1 US-08-276-887A-2	Sequence 2, Appli
C 40	15	1.9	894	1 US-08-178-708-7	Sequence 7, Appli
C 41	15	1.9	3987	1 US-07-688-352C-19	Sequence 19, Appli
C 42	15	1.9	78	1 US-08-281-702A-5	Sequence 5, Appli
C 43	15	1.9	9171	1 US-08-038-682-5	Sequence 6, Appli
C 44	15	1.9	9323	1 US-08-038-682-6	Sequence 6, Appli
C 45	15	1.9	2820	5 PCT-US94-05905-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1  
US-07-707-367-1  
; Sequence 1, Application US/07707367  
; Patent No. 5196316  
; GENERAL INFORMATION:  
; APPLICANT: Iwasaki, Yasuno  
; APPLICANT: Shimoi, Hiroko  
; APPLICANT: Suzuki, Kenji  
; APPLICANT: Ghisalba, Oreste  
; APPLICANT: Nishikawa, Yoshiki  
; APPLICANT: Kawahara, Takashi  
; APPLICANT: Kangawa, Kenji  
; TITLE OF INVENTION: No. 5196316el Enzyme and DNA Coding Therefor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/707,367  
; FILING DATE: 19910530  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 141678/90  
; FILING DATE: 01-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 210535/90  
; FILING DATE: 10-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 329911/90  
; FILING DATE: 30-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Villamizar, JoAnn  
; REGISTRATION NUMBER: 30,598  
; REFERENCE/DOCKET NUMBER: 4-18110/A/CGJ 44  
; TELECOMMUNICATION INFORMATION: /  
; TELEPHONE: (914)785-7120  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3383 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (recombinant)  
; ORIGINAL SOURCE:  
; ORGANISM: Xenopus laevis  
; INDIVIDUAL ISOLATE: DNA encoding protein AE-III,

INDIVIDUAL ISOLATE: precursor to PHL enzyme  
IMMEDIATE SOURCE:  
CLONE: PAE-III-202-4  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..2835  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1177..2145  
US-07-707-367-1

Query Match 2.1%; Score 17; DB 1; Length 3383;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAAAGTGGGACAA 351  
|||||  
1241 ACCTCAAAGTGGGACAA 1257

## RESULT 2

US-07-945-283-1  
Sequence 1, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Wesley, Ronald D.  
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
TITLE OF INVENTION: Involving The EP0 and LLT Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P  
REGISTRATION NUMBER: 27976  
TELEPHONE: 309-685-4011 ext.513  
TELEFAX: 309-685-4128

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8438 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 622..6495  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1099, "g")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1267, "t")

FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1381, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1566, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 2.1%; Score 17; DB 1; Length 8438;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 CCAAGATGGGGCGGCC 198  
|||||  
Db 7222 CCAAGATGGGGCGGCC 7238

## RESULT 3

US-08-261-206A-71/G  
Sequence 71, Application US/08261206A  
Patent No. 5574007  
GENERAL INFORMATION:  
APPLICANT: Zushi, Mitichitaka  
APPLICANT: Gomi, Komakazu  
APPLICANT: Yamamoto, Shuji  
APPLICANT: Suzuki, Koji  
APPLICANT: Matsuda, Akio  
TITLE OF INVENTION: A Polypeptide Capable of Interacting  
TITLE OF INVENTION: with Thrombin  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 N. Washington St.  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22046-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/261,206A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/740,492  
FILING DATE: 03-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30330  
REFERENCE/DOCKET NUMBER: 216-275P  
TELEPHONE: 703-241-1300  
TELEFAX: 703-241-2848  
TELEX: 248345

INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3306 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE:  
ORGANISM: Acremonium chrysogenum  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..3306

```
;; OTHER INFORMATION: /label= PKG_gene
;; OTHER INFORMATION: /note= "Nucleotide sequence of region A in Figure
;; OTHER INFORMATION: 59. The sequence is presented as Figure 61."
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 1252..1317
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 1463..1883
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 1948..2715
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: join(1252..1317, 1463..1883, 1948..2714)
US-08-261-206A-71

Query Match          2.1%; Score 17; DB 1; Length 3306;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 CTCCTCCCGCTCTGG 142
Db 209 CTCCTCCCGCTCTGG 193

RESULT 4
US-08-261-206A-75/c
; Sequence 75, Application US/08261206A
; Patent No. 5574007
; GENERAL INFORMATION:
; APPLICANT: Zushi, Mitichitaka
; APPLICANT: Goni, Komakazu
; APPLICANT: Yamamoto, Shuji
; APPLICANT: Suzuki, Koji
; APPLICANT: Matsuda, Akio
; TITLE OF INVENTION: A Polypeptide Capable of Interacting
; TITLE OF INVENTION: with Thrombin
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 N. Washington St.
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22046-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/261,206A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/740,492
; FILING DATE: 03-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30330
; REFERENCE/DOCKET NUMBER: 216-275P
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1251 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; ORGANISM: Acremonium chrysogenum
;; FEATURE:
;; NAME/KEY:
;; LOCATION: 1..1251
;; OTHER INFORMATION: /label= gene_sequence
;; OTHER INFORMATION: /note= "5' untranslated region of PKG gene from A.
;; OTHER INFORMATION: chrysogenum."
US-08-261-206A-75

Query Match          2.1%; Score 17; DB 1; Length 1251;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 CTCCTCCCGCTCTGG 142
Db 209 CTCCTCCCGCTCTGG 193

RESULT 5
US-08-759-945-1
; Sequence 1, Application US/08759945
; Patent No. 5834249
; GENERAL INFORMATION:
; APPLICANT: KAZUAKI, Furukawa
; APPLICANT: SUGIMURA, Keihiro
; APPLICANT: OHSUKE, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,945
; FILING DATE: 03-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/303,191
; FILING DATE: 08-SEP-1994
; APPLICATION NUMBER: JP 5-257861
; FILING DATE: 08-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 001560-227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2625 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2625
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 118..2625
US-08-759-945-1
```

Query Match 2.1%; Score 17; DB 3; Length 2625;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAGAGTGGGACAA 351  
|||||  
DB 1223 ACCTCAAGTGGGACAA 1239

RESULT 6  
US-08-994-719C-3  
; Sequence 3, Application US/08994719C  
; Patent No. 5908626  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Liming; Chang, Tse Wen  
; TITLE OF INVENTION: Hybrid with Interferon- and an  
; TITLE OF INVENTION: Immunoglobulin Fc Joined by a Peptide Liner  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Tanox Biosystems, Inc.  
; STREET: 10301 Stella Link Rd.  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch  
; COMPUTER: Adonics C142 SVGA  
; OPERATING SYSTEM: DOS 3.30  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/994,719C  
; FILING DATE: 12/19/1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/719,331  
; FILING DATE: 09/25/1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirabel, Eric P.  
; REGISTRATION NUMBER: 31,211  
; REFERENCE/DOCKET NUMBER: TX95-2AA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 664-2288  
; TELEFAX: (713) 664-8914  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1299 nucleic acids  
; TYPE: nucleic acid  
; STRANDEDNESS: double stranded  
; TOPOLOGY: linear  
US-08-994-719C-3

Query Match 2.1%; Score 17; DB 3; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GCGGAGCGGAGCGG 158  
|||||  
DB 580 GCGGAGCGGAGCGG 596

RESULT 7  
US-08-052-205-2/c  
; Sequence 2, Application US/08052205  
; Patent No. 5510259  
; GENERAL INFORMATION:  
; APPLICANT: SUGAMURA, KAZUO  
; APPLICANT: TAKESHITA, TOSHIKAZU  
; APPLICANT: ASAO, HIRONOBU  
; APPLICANT: NAKAMURA, MASATAKA

; APPLICANT: SHIMAMURA, TOSHIRO  
; APPLICANT: SUZUKI, MANABU  
; APPLICANT: HAMURO, JUNJI  
; TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/052,205  
; FILING DATE: 19930422  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 104947/1992  
; FILING DATE: 23-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5510259man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 10-615-0X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1393 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-052-205-2

Query Match 2.0%; Score 16; DB 1; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 CTACAGGACCTGGG 287  
|||||  
DB 1086 CTACAGGACCTGGG 1071

RESULT 8  
US-08-052-205-3/c  
; Sequence 3, Application US/08052205  
; Patent No. 5510259  
; GENERAL INFORMATION:  
; APPLICANT: SUGAMURA, KAZUO  
; APPLICANT: TAKESHITA, TOSHIKAZU  
; APPLICANT: ASAO, HIRONOBU  
; APPLICANT: NAKAMURA, MASATAKA  
; APPLICANT: SHIMAMURA, TOSHIRO  
; APPLICANT: SUZUKI, MANABU  
; TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington

STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/052,205  
FILING DATE: 19930422  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 104947/1992  
FILING DATE: 23-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Obolon, No. 5510259man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-615-0X  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1470 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 15..1121  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 81..1121  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 15..80  
US-08-052-205-3

Query Match 2.0%; Score 16; DB 1; Length 1470;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

272 CTACAGGACCCCTGGG 287  
|||||  
Db 1163 CTACAGGACCCCTGGG 1148

RESULT 9  
US-08-031-143B-68/c  
Sequence 68, Application US/08031143B  
Patent No. 5518880  
GENERAL INFORMATION:  
APPLICANT: LEONARD, WARREN J.; NOGUCHI, MASAYUKI;  
APPLICANT: MCBRIDE, O. WESLEY  
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND  
TREATMENT OF XSCID  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVE.  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORD PERFECT # 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/052,205  
FILING DATE: 12-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAM S. FEILER  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4061  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
TELEX: 421792

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORD PERFECT # 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/031,143B  
FILING DATE: 12-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAM S. FEILER  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4061  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1451  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: UNKNOWN  
MOLECULE TYPE:  
DESCRIPTION: OLIGONUCLEOTIDE  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
INDIVIDUAL ISOLATE: IL-2R  
US-08-031-143B-68

Query Match 2.0%; Score 16; DB 1; Length 1451;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 CTACAGGACCCCTGGG 287  
|||||  
Db 1163 CTACAGGACCCCTGGG 1148

RESULT 10  
US-08-031-143B-76/c  
Sequence 76, Application US/08031143B  
Patent No. 5518880  
GENERAL INFORMATION:  
APPLICANT: LEONARD, WARREN J.; NOGUCHI, MASAYUKI;  
APPLICANT: MCBRIDE, O. WESLEY  
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND  
TREATMENT OF XSCID  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVE.  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORD PERFECT # 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/031,143B  
FILING DATE: 12-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAM S. FEILER  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4061  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
TELEX: 421792

INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 301  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: UNKNOWN  
MOLECULE TYPE: OLIGONUCLEOTIDE  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
INDIVIDUAL ISOLATE: IL-2R  
US-08-031-143B-76

Query Match 2.0%; Score 16; DB 1; Length 301;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 CTACAGGACCTGGG 287  
|||||  
Db 301 CTACAGGACCTGGG 286

RESULT 11  
US-08-120-960-1/c  
Sequence 1, Application US/08120960  
Patent No. 5523225  
GENERAL INFORMATION:  
APPLICANT: KRAUS, JAN P  
TITLE OF INVENTION: DNA SEQUENCE ENCODING HUMAN  
TITLE OF INVENTION: CYSTATHIONINE B-SYNTHASE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DILWORTH & BARRESE  
STREET: 4350 LA JOLLA VILLAGE DRIVE, SUITE 300  
CITY: SAN DIEGO  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/120,960  
FILING DATE: 12-SEP-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: PEPPER PH.D., FREDERICK W.  
REGISTRATION NUMBER: 31,286  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4410  
TELEFAX: 619-453-2839  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2542 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 181..1834  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 181..1834  
US-08-120-960-1

Query Match 2.0%; Score 16; DB 1; Length 2542;

Best Local Similarity 100.0%; Pred. No. 36;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 TCCGCTCTCTGGCGG 146  
|||||  
Db 1826 TCCGCTCTCTGGCGG 1811

RESULT 12  
US-08-162-475A-1  
Sequence 1, Application US/08162475A  
Patent No. 5656474  
GENERAL INFORMATION:  
APPLICANT: Zohreh Tabaeizadeh  
TITLE OF INVENTION: A novel endochitinase gene  
TITLE OF INVENTION: induced by osmotic stress and abscisic acid isolated from the wild tomato Lycopersicon chilense Dun.  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1100 New York Avenue, N.W., 9th Floor  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 720 kb diskette  
COMPUTER: IBM PS/2, Model 30  
OPERATING SYSTEM: PC-DOS 3.30  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/162,475A  
FILING DATE: December 7, 1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, Paul E. Jr.  
REGISTRATION NUMBER: 32011  
REFERENCE/DOCKET NUMBER: PEW/3122/204351  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 966 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: CDNA to mRNA  
ORIGINAL SOURCE: Lycopersicon chilense  
US-08-162-475A-1

Query Match 2.0%; Score 16; DB 1; Length 966;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 565 TTGTCTCTTTTCTTG 580  
|||||  
Db 56 TTGTCTCTTTTCTTG 71

RESULT 13  
US-08-162-475A-3  
Sequence 3, Application US/08162475A  
Patent No. 5656474  
GENERAL INFORMATION:  
APPLICANT: Zohreh Tabaeizadeh  
TITLE OF INVENTION: A novel endochitinase gene  
TITLE OF INVENTION: induced by osmotic stress and abscisic acid isolated from the wild tomato Lycopersicon chilense Dun.  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

STREET: 1100 New York Avenue, N.W., 9th Floor  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 720 kb diskette  
COMPUTER: IBM PS/2, Model 30  
OPERATING SYSTEM: PC-DOS 3.30  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/162,475A  
FILING DATE: December 7, 1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, Paul E. Jr.  
REGISTRATION NUMBER: 32011  
REFERENCE/DOCKET NUMBER: PEW/3122/204351  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 935 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE: Lycopersicon chilense  
US-08-162-475A-3

Query Match 2.0%; Score 16; DB 1; Length 935;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 TTGTCCTCTTTTCTTG 580  
DB 27 TTGTCCTCTTTTCTTG 42

RESULT 14  
US-08-595-974-2/c  
Sequence 2, Application US/08595974  
Patent No. 5705608  
GENERAL INFORMATION:  
APPLICANT: SUGAMURA, KAZUO  
APPLICANT: TAKESHITA, TOSHIKAZU  
APPLICANT: ASAO, HIRONOBU  
APPLICANT: NAKAMURA, MASATAKA  
APPLICANT: SHIMAMURA, TOSHIRO  
APPLICANT: SUZUKI, MANABU  
APPLICANT: HAMURO, JUNJI  
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/595,974  
FILING DATE: 06-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/052,205  
FILING DATE: 22-APR-1993  
APPLICATION NUMBER: JP 104947/1992  
FILING DATE: 23-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5705608man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-615-0X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1393 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
ORGANISM: Homo sapiens  
US-08-595-974-2

Query Match 2.0%; Score 16; DB 2; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 CTACAGGACCTGGGG 287  
DB 1086 CTACAGGACCTGGGG 1071

RESULT 15  
US-08-595-974-3/c  
Sequence 3, Application US/08595974  
Patent No. 5705608  
GENERAL INFORMATION:  
APPLICANT: SUGAMURA, KAZUO  
APPLICANT: TAKESHITA, TOSHIKAZU  
APPLICANT: ASAO, HIRONOBU  
APPLICANT: NAKAMURA, MASATAKA  
APPLICANT: SHIMAMURA, TOSHIRO  
APPLICANT: SUZUKI, MANABU  
APPLICANT: HAMURO, JUNJI  
TITLE OF INVENTION: HUMAN IL-2 RECEPTOR GAMMA CHAIN MOLECULE  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/595,974  
FILING DATE: 06-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,205  
FILING DATE: 22-APR-1993  
APPLICATION NUMBER: JP 104947/1992  
FILING DATE: 23-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5705608man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-615-0X



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1470 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 15..1121  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 81..1121  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 15..80  
US-08-595-974-3

Query Match 2.08; Score 16; DB 2; Length 1470;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 CTACAGGACCTGGGG 287  
Db 1163 CTACAGGACCTGGGG 1148  
|||||

Search completed: February 10, 2000, 22:02:56  
Job time: 146 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 00:39:47 ; Search time 530.22 Seconds  
(without alignments)  
-4638.586 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTTAAAGGTC.....AAACGCAATTATATCCATAA 810

Scoring table: IDENTITY\_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.\*

Database size : 0

Number of hits that pass the threshold : 1642386

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_p11.\*
- 8: gb\_p12.\*
- 9: gb\_p13.\*
- 10: gb\_p14.\*
- 11: gb\_p15.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_on.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_p1.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*
- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*

50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	64.6	8.0	149592	33	AC005718	AC005718 Drosophil
C 2	64.6	8.0	188633	35	AC007175	AC007175 Drosophil
C 3	64.6	8.0	83204	46	AC004434	AC004434 Drosophil
C 4	64.6	8.0	132637	46	AC006092	AC006092 Drosophil
C 5	49.6	6.1	40912	35	CEL41D11	AF003740 Caenorhab
C 6	47.2	5.8	205157	44	AC016872	AC016872 Homo sapi
C 7	46.2	5.7	117674	32	DMR30C13	AL122025 Drosophil
C 8	46.2	5.7	1440	35	AF181623	AF181623 Drosophil
C 9	43.2	5.3	22333	34	CELC02F5	L14745 C. elegans
C 10	42.8	5.3	49649	16	CVY15035	Y15035 Cowpox viru
C 11	42.6	5.3	18753	2	AE001582	AE001582 Borrelia
C 12	41.6	5.1	88128	7	AB017089	AB017089 Arabidops
C 13	41.6	5.1	73384	40	AF178650	AF178650 Homo sapi
C 14	39.8	4.9	135301	16	BHV1CGEN	AJ004801 Bovine he
C 15	39.8	4.9	135301	16	BHV1CGEN	AJ004801 Bovine he
C 16	39.8	4.9	8113	16	HSBICP4A	L14320 Bovine herp
C 17	39.8	4.9	95456	33	AC002092	AC002092 Homo sapi
C 18	39.8	4.9	93713	40	AC006071	AC006071 Homo sapi
C 19	39.6	4.9	148051	32	AP000726	AP000726 Homo sapi
C 20	39.2	4.8	134977	11	AC002347	AC002347 Homo sapi
C 21	39	4.8	204929	45	AC009362	AC009362 Homo sapi
C 22	38.8	4.8	49173	42	AC011322	AC011322 Homo sapi
C 23	38.6	4.8	85835	8	ATAC005956	AC005956 Arabidops
C 24	38.6	4.8	158147	44	AC007926	AC007926 Trypanoso
C 25	38.4	4.7	123737	41	AF189001	AF189001 Homo sapi
C 26	38.2	4.7	219917	33	AC007612	AC007612 Homo sapi
C 27	38	4.7	1831	12	MUSNDRF	D83507 Mouse mRNA
C 28	38	4.7	12971	35	AF088979	AF088979 Dictyoste
C 29	38	4.7	179169	44	AC011946	AC011946 Homo sapi
C 30	37.8	4.7	14548	2	AF108766	AF108766 Rhodobact
C 31	37.8	4.7	126581	11	AC003957	AC003957 Homo sapi
C 32	37.8	4.7	35414	11	AC005328	AC005328 Homo sapi
C 33	37.8	4.7	43514	11	AC005545	AC005545 Homo sapi
C 34	37.8	4.7	170820	41	AC006286	AC006286 Homo sapi
C 35	37.8	4.7	194841	42	AC008155	AC008155 Homo sapi
C 36	37.8	4.7	165565	42	AC010775	AC010775 Homo sapi
C 37	37.6	4.6	100000	10	AP000502	AP000502 Homo sapi
C 38	37.6	4.6	210134	10	CNS0180V	AL109758 Human chr
C 39	37.4	4.6	126198	11	AC003089	AC003089 Human BAC
C 40	37.4	4.6	86894	11	AC003676	AC003676 Homo sapi
C 41	37.4	4.6	1582	40	AF151867	AF151867 Homo sapi
C 42	37.4	4.6	171681	40	HUAC002299	AC002299 Homo sapi
C 43	37.4	4.6	176384	42	AC009608	AC009608 Homo sapi
C 44	37.4	4.6	141193	44	AC015721	AC015721 Homo sapi
C 45	37.2	4.6	108803	11	HS550H1	AL035420 Human DNA

ALIGNMENTS

RESULT 1  
AC005718/c  
LOCUS  
DEFINITION  
AC005718 149592 bp DNA HTG  
Drosophila melanogaster chromosome 2 clone DS02336 (D440), map  
60C8-60D2 strain y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 68  
unordered pieces.  
AC005718  
VERSION  
AC005718.10 GI:5656710  
KEYWORDS  
HTGS\_PHASE1.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

# REFERENCE AUTHORS

1 (bases 1 to 149592)  
Celnikier, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenkov, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomontan, M.A., Mazda, P.,  
Moshrefi, M., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Sequencing of *Drosophila melanogaster*

# TITLE JOURNAL REFERENCE AUTHORS

2 (bases 1 to 149592)  
Celnikier, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,  
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,  
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,  
Lomontan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,  
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L.,  
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,  
Zieran, L.L. and Kimmel, B.E.

Direct Submission

# TITLE JOURNAL COMMENT

Submitted (26-SEP-1998) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Jul 30, 1999 this sequence version replaced gi:5630036.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bugreport@fruitfly.berkeley.edu](mailto:bugreport@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases. Pl library location:  
25-32.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 68 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 343: contig of 343 bp in length  
\* 344 423: gap of unknown length  
\* 424 1169: contig of 746 bp in length  
\* 1170 1249: gap of unknown length  
\* 1250 2094: contig of 845 bp in length  
\* 2095 2174: gap of unknown length  
\* 2175 3405: contig of 1231 bp in length  
\* 3406 3485: gap of unknown length  
\* 3486 4245: contig of 760 bp in length  
\* 4246 4325: gap of unknown length  
\* 4326 5054: contig of 729 bp in length  
\* 5055 5134: gap of unknown length  
\* 5135 6031: contig of 897 bp in length  
\* 6032 6111: gap of unknown length  
\* 6112 7163: contig of 1052 bp in length  
\* 7164 7243: gap of unknown length  
\* 7244 8013: contig of 770 bp in length  
\* 8014 8093: gap of unknown length  
\* 8094 8951: contig of 858 bp in length  
\* 8952 9031: gap of unknown length  
\* 9032 9815: contig of 784 bp in length  
\* 9816 9895: gap of unknown length  
\* 9896 10572: contig of 877 bp in length  
\* 10573 10652: gap of unknown length  
\* 10653 11442: contig of 790 bp in length  
\* 11443 11522: gap of unknown length  
\* 11523 12232: contig of 710 bp in length  
\* 12233 12312: gap of unknown length  
\* 12313 13278: contig of 966 bp in length  
\* 13279 13358: gap of unknown length  
\* 13359 14011: contig of 653 bp in length  
\* 14012 14091: gap of unknown length

14092 14897: contig of 806 bp in length  
\* 14898 14977: gap of unknown length  
\* 14978 15740: contig of 763 bp in length  
\* 15741 15820: gap of unknown length  
\* 15821 16793: contig of 973 bp in length  
\* 16794 16873: gap of unknown length  
\* 16874 17825: contig of 952 bp in length  
\* 17826 17905: gap of unknown length  
\* 17906 18911: contig of 1006 bp in length  
\* 18912 18991: gap of unknown length  
\* 18992 19720: contig of 729 bp in length  
\* 19721 19800: gap of unknown length  
\* 19801 20785: contig of 985 bp in length  
\* 20786 20865: gap of unknown length  
\* 20866 22392: contig of 1527 bp in length  
\* 22393 22472: gap of unknown length  
\* 22473 24290: contig of 1818 bp in length  
\* 24291 24370: gap of unknown length  
\* 24371 25694: contig of 1324 bp in length  
\* 25695 25774: gap of unknown length  
\* 25775 26731: contig of 957 bp in length  
\* 26732 26811: gap of unknown length  
\* 26812 27793: contig of 982 bp in length  
\* 27794 27873: gap of unknown length  
\* 27874 28161: contig of 1288 bp in length  
\* 28162 29241: gap of unknown length  
\* 29242 30491: contig of 1250 bp in length  
\* 30492 30571: gap of unknown length  
\* 30572 32263: contig of 1692 bp in length  
\* 32264 32343: gap of unknown length  
\* 32344 34053: contig of 1710 bp in length  
\* 34054 34133: gap of unknown length  
\* 34134 35477: contig of 1344 bp in length  
\* 35478 35557: gap of unknown length  
\* 35558 36683: contig of 1126 bp in length  
\* 36684 36763: gap of unknown length  
\* 36764 39017: contig of 2234 bp in length  
\* 39018 39097: gap of unknown length  
\* 39098 41095: contig of 1998 bp in length  
\* 41096 41175: gap of unknown length  
\* 41176 42262: contig of 1087 bp in length  
\* 42263 42342: gap of unknown length  
\* 42343 43644: contig of 1302 bp in length  
\* 43645 43724: gap of unknown length  
\* 43725 45000: contig of 1276 bp in length  
\* 45001 45080: gap of unknown length  
\* 45081 46626: contig of 1546 bp in length  
\* 46627 46706: gap of unknown length  
\* 46707 50898: contig of 4192 bp in length  
\* 50899 50978: gap of unknown length  
\* 50979 54635: contig of 3657 bp in length  
\* 54636 54715: gap of unknown length  
\* 54716 59775: contig of 5060 bp in length  
\* 59776 59855: gap of unknown length  
\* 59856 70727: contig of 10872 bp in length  
\* 70728 70807: gap of unknown length  
\* 70808 132766: contig of 61959 bp in length  
\* 132767 132846: gap of unknown length  
\* 132847 133588: contig of 742 bp in length  
\* 133589 133668: gap of unknown length  
\* 133669 134331: contig of 663 bp in length  
\* 134332 134411: gap of unknown length  
\* 134412 135076: contig of 665 bp in length  
\* 135077 135156: gap of unknown length  
\* 135157 135778: contig of 622 bp in length  
\* 135779 135858: gap of unknown length  
\* 135859 136505: contig of 647 bp in length  
\* 136506 136585: gap of unknown length  
\* 136586 137259: contig of 674 bp in length  
\* 137260 137339: gap of unknown length  
\* 137340 137991: contig of 652 bp in length  
\* 137992 138071: gap of unknown length  
\* 138072 138762: contig of 691 bp in length





Qy 650 GGTTCGTAATGGAGCGCTAATTGATTTCATTTTCAATGCACGATTTTGAC 709  
+ + + + +  
Db 5566 GCGCATGTTCCCTGGGCCAGCTGATTGACATCGTGCTGATAGCCTCGAGTCTGGGTC 5625

Qy 710 CTTCAGATGAAGTAGTACTATATAGATTACTATGGAACGAG 752  
+ + + + +  
Db 5626 CGCGGATGCTCGCGCTATGTATGATACCCTACTACGGAGCGG 5668

RESULT 5  
CEL41D1L/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CEL41D1L 40912 bp DNA INV 27-FEB-1998  
Caenorhabditis elegans cosmid C41D1L.  
AF003740  
AF003740.1 GI:2105490  
Caenorhabditis elegans strain=Bristol N2.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis:  
1 (bases 1 to 40912)  
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,  
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Faulstich, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,  
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
Wilkinson-Sporat, J. and Wohldman, P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
Nature 368 (6466), 32-38 (1994)  
94150718  
2 (bases 1 to 40912)  
Gattung, S and Maggi, L.  
The sequence of C. elegans cosmid C41D1L  
Unpublished (1998)  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RO, England  
e-mail: rw@nemato.de.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is cllp9, 3100 bp overlap. Actual start of this  
cosmid is at base position 1 of CELC41D1L; actual end is at 40912  
of CELC41D1L

NOTES:

Coding sequences below are predicted from computer analysis, using  
the program Genefinder(P. Green and L. Hillier. ms in preparation)



```
Ds 25983 AACGTAATTTCTACAGCGAATTCATGTAACGTGTCATCTGCATACAGTTGGACAAAA 25924
QY 557 CAGTCGATGTCCTTTTCTTTGGATGGTTGGGACAGATCGATTTTACCTTTGGATACC 616
Ds 25923 CAATGATTTCTCAGTTGTCGCGGGTTTGGAGCTGATCTTTCTACTTGGGACTCT 25864
QY 617 CTGCTTTGGTTTGAAGATTTTGCACCTAGGCTTTTGTGAAATGGGAGCCTAATTG 676
Ds 25863 GGAATCTGCAATTTGAAACATTTTCAGTTTCGTTGGTCTCTTGTGTGGACTCTTGTGG 25804
QY 677 ATTTCATCTTATTTCATGCAGATGTTTGGACCTTCAGATGGAAGTAGTTACATTATAG 736
Ds 25803 ACCTTGTTCGATCGCAGTTGGATGATTAACCTTACGATGATCCATGATATTGAT 25744
QY 737 ATTACTATGGAACCACTTACAA 760
Ds 25743 TTTTITTTTAAATTTCTTTTCCAA 25720

RESULT 6
16672
US
DEFINITION AC016672 205157 bp DNA HTG 04-DEC-1999
Homo sapiens clone RP11-14C10, WORKING DRAFT SEQUENCE, 1 unordered
pieces.
ACCESSION AC016672
VERSION AC016672.1 GI:6524410
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 205157)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205157)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H_NH0014C10.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 205157: contig of 205157 bp in length.
FEATURES
Location/Qualifiers
source
1..205157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-14C10"

BASE COUNT 53433 a 46205 c 47212 g 58163 t 144 others
ORIGIN

Query Match 5.8%; Score 47.2; DB 44; Length 205157;
Best Local Similarity 56.4%; Pred. No. 0.061;
Matches 88; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 577 CTGGAATGGTTGGACAGACAGATGATTTTACCTTGGATACCCCTGCTTTGGTTGTTAAAG 636
Ds 15774 CTCGGTGGGTTTGGACAGACACCTTTCTACCTGGCCAGTGGGGAAGCCCTCGGCAAG 15833
QY 637 TTTTGACATGTAGGCTTTTGTGAAATGGGAGCCTTAATTCATTCTATTTCATG 696
Ds 15834 CTCTTCAGCTTCGTCGTCCTGGGAATATGACGCTGATAGACGTCCTGCTCATTTGGATT 15893
QY 697 CAGATGTTGGACCTTCAGATGGAAGTAGTTACATT 732
```

Ds 15894 GGCTATGTTGACCAGCAGATGGCTCTTTGTACATT 15929

# RESULT 7

DMR30C13

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This is a 'working draft' sequence. It currently

consists of 153 contigs. The true order of the pieces is not known

and their order in this sequence record is arbitrary. Gaps between

the contigs are represented as runs of N, but the exact sizes of

the gaps are unknown. This record will be updated with the finished

sequence. 1 513: contig of 513 in length

514 613: gap of unknown length

514 1380: contig of 767 in length

1381 1480: gap of unknown length

1481 1995: contig of 515 in length

1996 2095: gap of unknown length

2096 2862: contig of 767 in length

2863 2962: gap of unknown length

2963 3754: contig of 792 in length

3755 3854: gap of unknown length

3855 4657: contig of 803 in length

4658 4757: gap of unknown length

4758 5330: contig of 573 in length

5331 5932: contig of 502 in length

5933 6032: gap of unknown length

6033 6474: contig of 442 in length

6475 6574: gap of unknown length

6575 7188: contig of 614 in length

7189 7288: gap of unknown length

7289 7513: contig of 225 in length

7514 7613: gap of unknown length

7614 8288: contig of 655 in length

8289 8368: gap of unknown length

8369 9346: contig of 978 in length

9347 9447: gap of unknown length

9447 10066: contig of 620 in length

10067 10166: gap of unknown length

10167 10814: contig of 648 in length

10815 10914: gap of unknown length

10915 11203: contig of 289 in length

11204 11303: gap of unknown length

11304 11941: contig of 638 in length

11942 12041: gap of unknown length

12042 12665: contig of 624 in length

12666 12765: gap of unknown length

12766 13410: contig of 645 in length

13411 13510: gap of unknown length

13511 13996: contig of 486 in length



13997 14096: gap of unknown length  
14097 14677: contig of 581 in length  
14678 14777: gap of unknown length  
14778 15318: contig of 541 in length  
15319 15418: gap of unknown length  
15419 16047: contig of 629 in length  
16048 16477: gap of unknown length  
16478 16888: contig of 641 in length  
16889 17502: contig of 614 in length  
17503 17602: gap of unknown length  
17604 18281: contig of 679 in length  
18282 18381: gap of unknown length  
18383 18819: contig of 438 in length  
18820 18919: gap of unknown length  
18920 19754: contig of 835 in length  
19755 19854: gap of unknown length  
19855 20599: contig of 845 in length  
20700 20799: gap of unknown length  
20800 21183: contig of 384 in length  
21184 21283: gap of unknown length  
21284 22021: contig of 738 in length  
22022 22121: gap of unknown length  
22122 23187: contig of 1066 in length  
23188 23287: gap of unknown length  
23289 23945: contig of 658 in length  
23946 24045: gap of unknown length  
24046 24616: contig of 571 in length  
24617 24716: gap of unknown length  
24717 25676: contig of 960 in length  
25677 25776: gap of unknown length  
25778 26190: contig of 414 in length  
26191 26290: gap of unknown length  
26291 27292: contig of 1002 in length  
27293 27392: gap of unknown length  
27394 28107: contig of 715 in length  
28108 28207: gap of unknown length  
28209 28729: contig of 522 in length  
28730 28829: gap of unknown length  
28830 29576: contig of 1147 in length  
29577 30076: gap of unknown length  
30077 30928: contig of 852 in length  
30929 31028: gap of unknown length  
31029 31903: contig of 875 in length  
31904 32003: gap of unknown length  
32004 32418: contig of 415 in length  
32419 32518: gap of unknown length  
32519 33132: contig of 614 in length  
33133 33232: gap of unknown length  
33233 33823: contig of 591 in length  
33824 33923: gap of unknown length  
33924 34827: contig of 904 in length  
34828 34927: gap of unknown length  
34929 35934: contig of 1007 in length  
35935 36034: gap of unknown length  
36035 36962: contig of 928 in length  
36963 37062: gap of unknown length  
37064 37679: contig of 617 in length  
37680 37779: gap of unknown length  
37780 38484: contig of 705 in length  
38485 39541: contig of 957 in length  
39542 39642: gap of unknown length  
39643 40517: contig of 876 in length  
40518 40617: gap of unknown length  
40618 41307: contig of 690 in length  
41308 41407: gap of unknown length  
41408 42011: contig of 604 in length  
42012 42111: gap of unknown length  
42112 42828: contig of 717 in length  
42829 43593: contig of 665 in length  
43594 43693: gap of unknown length

43594 44391: contig of 598 in length  
44392 44491: gap of unknown length  
44492 45147: contig of 656 in length  
45148 45247: gap of unknown length  
45249 46798: contig of 1551 in length  
46799 46898: gap of unknown length  
46900 47830: contig of 932 in length  
47831 47930: gap of unknown length  
47932 48883: contig of 953 in length  
48884 49341: gap of unknown length  
49342 50441: gap of unknown length  
50442 51797: contig of 1356 in length  
51798 52237: contig of 340 in length  
52238 52337: gap of unknown length  
52339 53710: contig of 1373 in length  
53711 53810: gap of unknown length  
53812 54430: contig of 620 in length  
54431 54530: gap of unknown length  
54532 55595: contig of 1065 in length  
55596 56555: contig of 860 in length  
56556 56655: gap of unknown length  
56656 57909: contig of 1254 in length  
57910 58009: gap of unknown length  
58010 59436: contig of 1427 in length  
59437 59536: gap of unknown length  
59537 60724: contig of 1188 in length  
60725 60824: gap of unknown length  
60826 62193: contig of 1369 in length  
62194 62293: gap of unknown length  
62294 63059: contig of 776 in length  
63060 63159: gap of unknown length  
63160 64233: contig of 1064 in length  
64234 64333: gap of unknown length  
64334 64782: contig of 449 in length  
64783 64882: gap of unknown length  
64883 66244: contig of 1362 in length  
66245 66344: gap of unknown length  
66346 67507: contig of 1163 in length  
67508 67607: gap of unknown length  
67609 68877: contig of 1270 in length  
68878 68977: gap of unknown length  
68979 70692: contig of 1715 in length  
70693 70792: gap of unknown length  
70794 71673: contig of 881 in length  
71674 71773: gap of unknown length  
71774 72809: contig of 1036 in length  
72810 74298: contig of 1389 in length  
74299 74398: gap of unknown length  
74399 76722: contig of 2324 in length  
76723 76822: gap of unknown length  
76824 79379: contig of 2557 in length  
79380 79479: gap of unknown length  
79480 79862: contig of 383 in length  
79863 79962: gap of unknown length  
79964 80475: contig of 513 in length  
80476 80575: gap of unknown length

Query Match

Best Local Similarity 5.7%; Score 46.2; DB 32; Length 117674;

Matches 87; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 577 CTTGGATGTTGGGACGATCGATTTTACCTTGGATACCCCTGCTTGGTTGTTTAAAG 636

Db 4279 CTAGGGGCTTTGGAGCCGATCGATCTACTTGGCCACTGCCAGAGGGGATGGAAG 4338

Qy 637 TTTTCACGTAGGGTTTGTGGAAATGGGACGCTAAATGATTCATTCTTATTTCAATG 696

Db 4339 CTGTTACGCTTTGGCGGCTCGCGCTCTGGACCATCATCGATGTCCTGCTCATCTCGATG 4398

```
QY 697 CAGATTGTTGGACCTTCAGATGGAAGTAGTTACAT 731
|| | | | | | | | | | | | | | | | | | | |
Db 4399 CATTATTGGGGCCGCGGATGCTCACTTTACAT 4433

RESULT 8
AF181623 1440 bp mRNA INV 16-SEP-1999
LOCUS Drosophila melanogaster BCDNA.GH02340 (BCDNA.GH02340) mRNA,
DEFINITION complete cds.
ACCESSION AF181623
VERSION AF181623.1 GI:5901801
KEYWORDS FLI_CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1440)
Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
Snir,E., Svirska,R.R., Weinburg,T. and Celniker,S.E.
Full Length Drosophila melanogaster cDNA sequence
Unpublished
2 (bases 1 to 1440)
Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
Snir,E., Svirska,R.R., Weinburg,T. and Celniker,S.E.
Direct Submission
Submitted (27-AUG-1999) Berkeley Drosophila Genome Project,
University of California Berkeley, 539 Life Sciences Addition
#3200, Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
University of California Berkeley
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
TUES Location/Qualifiers
source
1. .1440
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH02974"
1. .1440
/organism="BcdNA.GH02974"
/feature="motif=(desc:'C_TYPELECTIN_1', acc:'PS00615',
count:'1', method:'PPSEARCH', dbase:'PROSITE')",
276..1130
/feature="BcdNA.GH02974"
/codon_start=1
/product="BcdNA.GH02974"
/protein_id="AAD55409.1"
/db_xref="GI:5901802"
/db_xref="FLYBASE:FBcd0002974"
/translat="MRLQRCQIVNMRSAIVLIMFVLTGRINSETASGGNQMDLSDS
RGDKRDNNSANGNANDNEVPLVSWAKGGGAGGLDINITYSSSSSSSS
NNNNMLCPYKETPCDRLOPCRCNTYHGGCIYGRDLNVTCTVNNVOCIGERSFOR
QMRCYCYQTEMWQSCQSSNSATDKLFTNCTVHHDVLCGNRSTRNLRCNWT
QGYRWSTALLISLIUGGFGADRFYLGHWQEGIGKLFSGGLGVTWIIIDVLLISMHYLG
PADGSLYI"
BASE COUNT 422 a 325 c 329 g 364 t
ORIGIN
```

```
Query Match 5.7%; Score 46.2; DB 35; Length 1440;
Best Local Similarity 56.1%; Pred. No. 0.046;
Matches 87; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 577 CTGTGGATGGTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTGGGTTGTTAAAG 636
|| | | | | | | | | | | | | | | | | | | |
Db 972 CTAGGGGCTTTGGAGCGGATCGATTTTACTTTGGCCCACTGGCAGGAAGGATTGGAAG 1031
|| | | | | | | | | | | | | | | | | | | |
QY 637 TTTTGCAGCTAGGCTTTTGTGAATTGGGAGCCCTAAATTTTCAATCTTATTTCAATG 696
|| | | | | | | | | | | | | | | | | | | |
Db 1032 CTGTTTCACTTTGGCGGCTCGGCTGGACCATCATCGATGCTCTCATCTCGATG 1091
|| | | | | | | | | | | | | | | | | | | |
QY 697 CAGATTGTTGGACCTTCAGATGGAAGTAGTTACAT 731
|| | | | | | | | | | | | | | | | | | | |
Db 1092 CATTATTGGGGCCGCGGATGCTCACTTTACAT 1126
|| | | | | | | | | | | | | | | | | | | |

RESULT 9
CELCO2F5 22333 bp DNA INV 22-OCT-1993
LOCUS C. elegans cosmid C02F5.
DEFINITION L14745.L18807
ACCESSION L14745.L18807
VERSION L14745.1 GI:289607
KEYWORDS GRR1 gene; GTP-binding protein; glucose repression; proteasome
component C5.
SOURCE Caenorhabditis elegans (strain Bristol N2) hermaphrodite mixed
whole animal DNA.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 22333)
Sulston,J., Du,Z., Thomas,K., Willson,R., Hillier,L., Staden,R.,
Halloran,N., Green,P., Thierry-Mieg,J., Qiu,L., Dear,S.,
Coulson,A., Craxton,M., Durbin,R., Berks,M., Metzstein,M.,
Hawkins,T., Ainscough,R. and Waterston,R.
The C. elegans genome sequencing project: A beginning
Nature 356, 37-41 (1992)
92168156
2 (bases 1 to 22333)
Anderson,K.
Sequence of the C. elegans cosmid C02F5
Unpublished (1993)
Submitted by:
Nematode Sequencing Project
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
MRC Laboratory of Molecular Biology,
Cambridge CB2 2QH, England
e-mail: rwnematode.wustl.edu and jes@cele.mrc-lmba.cam.ac.uk
NOTE:
Coding sequences below are predicted from computer analysis using
the program Genefinder (P. Green and L. Hillier, in preparation)
Neighboring cosmid information:
Bases 1 to 800 of this entry correspond to bases 26944 to 27743 of
GenBank entry CELC30A5 (L10990). Bases 22034 to 22333 of this
entry
correspond to bases 1 to 300 of GenBank CELF09G8 (L11247). The
sequence
of C. elegans cosmid C02F5 begins at position 18944 of GenBank
CELCO30A5
(L10990), continues through this entry and ends at position 15818
of
GenBank CELF09G8 (L11247). The sequence of C. elegans cosmid F09G8
begins at position 22034 of this entry.
Location/Qualifiers
source
1. .22333
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/dev_stage="mixed"
/sex="hermaphrodite"
/tissue_type="whole animal"
/map="3"
```

```

CDS
join(8..124,175..257,305..594,742..943,993..1493,
2247..2423)
/standard_name="C02F5.7 protein"
/notes="homology with glucose induced repressor, GRRL1;
putative"
/codon_start=1
/protein_id="AAA27922.1"
/db_xref="GI:289614"
/translation="METAERIISAAASRRRAKRLQAQAKHTPVIQAKONQMYLIT
TLSPADVNDLNLNRPKEVLLKVFSLDTKALCRSAQVCRSWIILALDGSNMQRVCL
FTFQDVKTAVENLARCGFLKELSLKGCENVHDSALRTFTSRCPNLEHLSLYRCK
RYTASCEENLGRVCHKLNYLNLENCSITDRANKYIGDGCNLSYLNISWCDALQDRG
QVILSCKSLDILILRGCEGLTENFGSVEAHMGAIKKLNLCOCFQLTIDTVGNIAN
GATALEYLCSNQNQISDLSVLSGQSHNLKVLSELGCTLLGDMGFIPLRAGCRQLE
RLMDEDCSLSDHTINSLANNCTALRESHCLELITDESIONLASHKHRETFLNLYELD
NCPLTDSLSHLRCHKALRIDLYDQNTYSKEAIRFQVVDVAVLSYDCIYVPCSTR
STENWYKLTIKTRKSDINRAPCAHVRHAQLWYSIKWLKF"
175..257
/notes="putative"
/number=9
305..694
/notes="putative"
/number=10
742..943
/notes="putative"
/number=11
993..1493
/notes="putative"
/number=12
misc_feature
2786
/notes="possible trans-spliced leader site for C02F5.6"
join(2793..2834,3022..3473,3528..3686,3737..4091,
4139..4252,4326..4463)
/standard_name="C02F5.6 protein"
/notes="putative"
/codon_start=1
/protein_id="AAA27921.1"
/db_xref="GI:289613"
/translation="MAHTSDGWPYDQNTQYVEAEQLEIALLEPLDRILETANVEFF
RPAFKNHQDNDNRKKNNDNEWRDSIYNATDESDDTDHEQRKNFFQPPLOVORNSF
VKNTLMFEKSSQIDISRLAVMCGGEMSLKGEYLGSGGTINLVSDIDEPSLSIG
QQLGKAESIFLRKHLERNAELIYAVETGLPVMRSYVGDILEPDRHFAVDVAISMEV
VEHIPUPNAKFFVENLGTLMPIRIFETPSHNEYNAVFGMEPGFGRHGDHFEMNRKE
FSNWLELSIRFPHYQIDPPHYIGMTGRYENLSGSAQAACRLQVDLNTLTPQSVTPY
EMVGHLPCLRGLSLIAYNLVKEAFLDWLEKLELQEHPRTDGYSPIYFNVQNLHL
KAPVSFALTIDEXVAMQVSSVIGTVEF"
3022..3473
/notes="putative"
/number=13
3528..3686
/notes="putative"
/number=14
3737..4091
/notes="putative"
/number=15
4139..4252
/notes="putative"
/number=16
join(6347..6436,6481..6795,6876..6992)
/standard_name="C02F5.5 protein"
/notes="putative"
/codon_start=1
/protein_id="AAA27920.1"
/db_xref="GI:289612"
/translation="MSMTEQIPPEYKYVGRCAVLACSSFELMMVLGSLGDSNV
LAKFLVFTGSAAGVFNQIDGRDEIKKITGSDNETRGKAAALVGIIPALAAV
FVFLCVSGHAFFSLFVLVHLASIGQLGIEAYEVRNGFCIKAPENKKAVADSYDQNTQ
ILNMVQDSVFDKK"
6481..6795
/notes="putative"
/number=17
complement(join(8274..8477,8518..8583,8631..8897,
9041..9178,9473..9532))
/standard_name="C02F5.8 protein"
/notes="putative"
/codon_start=1
/protein_id="AAA27923.1"
/db_xref="GI:289615"
/translation="MATWKFIRSVLFFELDLAMLLAALALIAVGFWMGYDSSFDTLK
NVIYIDDPKSLADAKNIRVWLIVFWSIIGLSGAVVTAVIGWISSVWPKRGKGMFI
TENSLLIYVLSGECGVAVLVRNLSHDNTSLDAMVYTTNSVNDLKTIQKYNCCG
IYSLFNVMYCGPMSCPKCDVAVFDSVDNTMMISGILLVILQITAILLPVPLI
SRKTKYKSYSPRVTLADITEDTRF"
8518..8583
/notes="complement; putative"
/number=6
8631..8897
/notes="complement; putative"
/number=7
9041..9178
/notes="complement; putative"
/number=8
/notes="possible trans-spliced leader site for C02F5.8"
complement(join(11079..11495,11809..12165,12215..12369,
13369..13474,13523..13916,13962..14227))
/standard_name="C02F5.9 protein"
/notes="homology with proteasome component C5; coded for by
C. elegans cDNAs GenBank:M80127 and T01498; putative"
/codon_start=1
/protein_id="AAA27924.1"
/db_xref="GI:289616"
/translation="MISFTGITAVANEMAMFQAMKEVAHPEWSSROIERQRW
NPYMEGGSTCAISGENFAIVASDTRMTQNDINILTRDAEKIOLINDNIIILTSFYG
DVLQKYLQSLRHKYRFDYRSDMSVLCALLSRNLYRFFYTGAILAGIDHGH
KGAVSYDPIGICIERLGSASGAAPMIIPFLDCOIGHVTLSEGYTERPELTLDATSL
MKDSFGAAREIEISTGDKHLVLAEGKPVVVKFLPLOSISAQSONVLKPKKSTSAE
LGFAPQLNLVDTLGLKPRDSAMSGKRPASVLKGLFDWDNDTANLDIPGTSIDPESK
PSSALKVPKEVLDSEIELDQTSSSSFLNDSQENQERNILEDLRLQOMFTPSK
IKNEEASLLKCGTNGKELESSESHANYITNFCGGLKYNRWYKYLRLDHALTH
SNRIKPCKLCGFEKTNVRMRSHYAKAHPNERVEGYGMKALVSGDSSRIGDQVGDID
QOVSDELKELWNCFSESILHVGQASGVEGEKYRMTKRKRLDREAMNSHNEF"
11809..12165
/notes="complement; putative"
/number=2
12215..12369
/notes="complement; putative"
/number=3
13369..13474
/notes="complement; putative"
/number=4
13523..13916
/notes="complement; putative"
/number=5
join(14580..14717,14761..14894,14937..15120,15330..15412,
15465..15612,15661..15759)
/standard_name="C02F5.4 protein"
/notes="coded for by C. elegans cDNA GenBank:CELI6D4
(M89109); putative"
/codon_start=1
/protein_id="AAA27919.1"
/db_xref="GI:289611"
/translation="MADTEQTLRQKTLNLSNHPSSIOTTSAWLQNHSNRLIIRVW
LKTVKETHGSKIVNLLYVANDVSONARKTCPOKDEFFPAIESFRHAEIKUNAKEV
EHAIGKLNWKKDRIETPSCOKRLHEVQOVLKSGSPTPAVANKHGKSVRSQEA
QSLVQNEALPMLEEYVKRLKNETNERETLESNMLIENVRMSIEHHEKLCREVKR
REDRIKADLLEVEKTFESLPDLAAEMPNAPLTLEALFEKRRK"
14761..14894
/notes="putative"
/number=18
14937..15120
/notes="putative"
/number=19
15330..15412
/notes="putative"
/number=20

```



```
gene
7458..8969
/ gene="B2R"
CDS
7458..8969
/ gene="B2R"
/codon_start=1
/protein_id="CAA75278.1"
/db_xref="GI:3097023"
/db_xref="SPTREMBL:O72739"
/translation="MAMFYALGQYDENLHAFPGISSTVANDVRKYSVSVYNNKYD
IVDKYMWYSEVKNRYIGALLPMFECNEYLQIDPIHDQEGNOISITIRKKNYAL
SGIGESLDLCEGVGTHHVLETGNVYGVQHDYSTIREKAKEMSTLSPGPIIDVH
VWIGDCICVQADVNDVHVKELMKREFKGAQLQIPENLVKLGENDTENLSTASAILN
SGGCTEIVTSQEDRVNHLVLMKRESIRHMSVYVDHEDIVNGKCCYVHMSSQNPI
PSVTKNLKMTKWSCLQMDMSALEYLSKESGSRPRPEMOEFPYDGSFANTEGG
ERLAEEFNSELOAGSVKFGNSINVKHTSVSAKOLRIRQOLPESLVDVTEG
YLFIQVNNTHKVIQFVGHDLRLVENDIEKHKRLRVVHFCEKKEDIKYACRFKV
YKPGDETSYVCAIKVERCCAVFADWPESWYMDTNGIKKYSPEWVSHKF"
9194..10870
/ gene="B3R"
9194..10870
/ gene="B3R"
/codon_start=1
/protein_id="CAA75279.1"
/db_xref="GI:3097024"
/db_xref="SPTREMBL:O72740"
/translation="MDFFKKELDWSYLSHYIARACSNSTSHIIQENVLIRTEYK
LGTIVDFLSRWPNVLEHILEYGENILHIYMDANANIIFFLDRVLNKNKGSFIH
LGSSSINIKYVQVNVNDPNRIKMLENGRRTRHFLSYILDVNIYISILNHG
FYDAEDSYCTLLHRSIYHYKKSSESYNELIKLLNGSDVDKDKTYGNTFILLC
KHIDNAELFEIENLANSIDVPDTPHYVSCRNYDKYFVKSLSKGANVARNRF
GTPFFCGIHLGSLIKLYLESDELEIDNEHIVRHLLIIFDAVESLDYLLSRGVIDIN
IRIYISYISIPVSNAYNATLVYLNENRGDEFITTSGGCTISEAVANNKIIEMEYL
LSRPSLKMIPDAIATRHKNADLLKMCIKYACTMDYDILIDVQSLOQYKWIYL
CFDEIDMKRCYIKNKTQVLCIKINDINTLMRYGKHPKFCVFCARLDVYGSVRNII
ASRYRORLISLSKLLDAGDKWSCFPNEPKYIKYLENFENDNELSTYLLKIL"
10974..11927
/ gene="B4R"
10974..11927
/ gene="B4R"
/codon_start=1
/protein_id="CAA75280.1"
/db_xref="GI:3097025"
/db_xref="SPTREMBL:O72741"
/translation="MKTISVTLCLVLPVAVYVCTVPTMNAKLSTSTERSFNDKQKV
TFTCDGYHSLDPAVCEVDKWKYENPCPKMCTSDYVSELYDKPLYEVASTMTLSCN
GETKYRCEKNGTNSWNTVTCNACQPLQLEHSGCQPVKEKYSFGEYMTINCDVG
YEVIQASYISCTANSWNVIPSCQKCDIPSLNGLISGSTFESIGGVHLKSKSGFTLT
GPSFSCIDGKNRPILPTCVRSNEEFLVDGDDDETDLSKISKDQVVOYEQIESLEA
TYHIIIVALTIMGVIFLISIIIVLVCSDKNNDQYKFHKLIP"
12023..12574
/ gene="B5R"
12023..12574
/ gene="B5R"
/codon_start=1
/protein_id="CAA75281.1"
/db_xref="GI:3097026"
/db_xref="SPTREMBL:O72742"
/translation="MSSSDVDIVDAVRAFLLRHYDKRITVYGRSNAILLNIVRLFT
RCVAPDDIVRTPMPSRVKQWVIDFLNDIMKEHDVSVSGIGILFMEMFFDYKNK
NSNSNSKNSINQLMIDIINSVSIILANERYSAFNDDGIYIRMTMDKLYGYASL
TTIGMITGGVYCYLLMHLVSLYK"
12613..13161
/ gene="B6R"
12613..13161
/ gene="B6R"
/codon_start=1
/protein_id="CAA75282.1"
/db_xref="GI:3097027"
/db_xref="SPTREMBL:O72743"
/translation="MYKNLITFLVIGAVASYSNNEYTPPNKLSVKLYIDGVNIIENS
YTDNNELVNFREYTIISITESCVDGFDSDIDVIDYKIIDMYITIDSITQRKHT
CKTSLKLSKYPYKIHXYIDGDERQYSITAEGKYKIGKIYEISMMNDDTLRLKKTLLK
IGSTYIFDRHGSNTYYSKYDF"
```

```
gene
13239..14054
/ gene="B7R"
Query Match
Best Local Similarity 5.3%; Score 42.8; DB 16; Length 49649;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 652 TTTTGTGAATGGGAGCCTAATTGATTCATCTTATTTCATGAGATGTTGGACCT 711
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38470 TTCTCTGTAATGATGATTTAAACATGATTAATTGTTCTTTTATGAATTTGTTTCAT 38411
QY 712 TCAGATGGAAGTAGTACATATTAGATTACTATGGAACACAGACTTACACAGACTGAGTATT 771
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38410 TAAAGTGCATAAGTATTATATATCTTACGTAGTAGTATCAGTTGTAAGATATGGAATAAT 38351
QY 772 ACTAATGAACATTTAGAAAAACGCAATATATATCCATA 809
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38350 ACTAAGCAATGATTGATTACGGATAAAGATATATCTA 38313

RESULT 11
LOCUS AE001582/c
DEFINITION Borrelia burgdorferi plasmid lp21, complete plasmid sequence.
ACCESSION AE001582
VERSION AE001582.1 GI:63822381
KEYWORDS
SOURCE Lyme disease spirochete.
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group.
REFERENCE
1 (bases 1 to 18753)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
A bacterial genome in flux: The twelve linear and nine circular
extrachromosomal DNAs in an infectious isolate of the Lyme disease
spirochete Borrelia burgdorferi
Mol. Microbiol. (1999) In press
2 (bases 1 to 18753)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
Direct Submision
Submitted (11-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
FEATURES
source
1..18753
/organism="Borrelia burgdorferi"
/db_xref="taxon:139"
/plasmid="lp21"
184..618
/ gene="BBU01"
184..618
/ gene="BBU01"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF07699.1"
/db_xref="GI:6382388"
/translation="MENQYFENLFFIKFIYLYIMVKSVEQOEKGDHTCRILLI
LTEIFENLILENSQTLKKYNENLKNLPPSNIMKYLNLQLEKIKIANK
YFKNDQSLIYKLYNTDEKLIWKLIELFYKELQFIQNTTT"
746..1114
/ gene="BBU02"
746..1114
/ gene="BBU02"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
```

gene /product="hypothetical protein"  
/protein\_id="AAF07700.1"  
/db\_xref="GI:6382389"  
/translation="MDGVINDTLVARMKKQIKFNKKILILVKTLDHMKELLYSANK  
TLYVLQNNENEALAKYQLRVNYKTLLEYLEILKPNKPKILKPTNKENESFIGLY  
TLVPLEDCCKIKYNSHNI"  
complement(1238..1357)  
/gene="BBU03"  
CDS complement(1238..1357)  
/gene="BBU03"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF07701.1"  
/db\_xref="GI:6382390"  
/translation="MVYKLVRLYARIDVIISNEVGLREISLKEIVLSPIHLV"  
1486..2628  
/gene="BBU04"  
CDS 1486..2628  
/gene="BBU04"  
/note="similar to PID:1174341 PID:1174349 GB:AE001584  
percent identity: 81.23; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF07693.1"  
/db\_xref="GI:6382382"  
/translation="MLISFVKESIPMHKIKITNNPNRCYNKQVKKLVILSTICYLN  
KTHKKYQKTLTYFNENLRNGQITSLRTMQRYIYRLOKEIKVTTNYQHMVNSG  
TEIYKLYNQDCYHKINOHFKKKEKTRFQNRVNYFNKNSDKMGVSQVCESSNSK  
NNKEKINIEKYOVNYFNKCNFSCKEILSILLNVDKDMIKIKTKTRDIK  
AKNIYPPKSCKEQKELKILCINTOKELEKSGYSEQLETNFKQIYENKYPHF  
IENHKYSDLSIARKKLKESIERKENSQDKNLRITNFILIEQLKKTNIETLKP  
IIRKYLNNKKIKYKFRYYISELLEIINGKHYSNLAKFRKKSNG"  
2868..3656  
/gene="BBU05"  
CDS 2868..3656  
/gene="BBU05"  
/note="similar to GB:D26185 SP:P37522 PID:467381  
PID:580906 GB:AL009126 percent identity: 55.98; identified  
by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="plasmid partition protein, putative"  
/protein\_id="AAF07694.1"  
/db\_xref="GI:6382383"  
/translation="MNKKPNLITASLGGVGKSTLTILFSYLLKDLGKKILIVDMDP  
QNSTYSFNVHVEIYNTYMLKGGVSNVCYCKINDHIFILPAHEILENFLEPI  
DYKIVLEIREQNIISAYEPDYLDDTPPHRDLVKKNALNITDHIIPQVERWSIES  
FTILMNMIENQILKKITITSIVENQFKNRDLKDIERLLYKYSYVKGKIHFSN  
GIKVFINELEPPSEESYKKAETLNLILYINHNKRYFKNK"  
14633..15235  
/gene="BBU06"  
CDS 14633..15235  
/gene="BBU06"  
/note="similar to GB:U03641 PID:458218 percent identity:  
61.64; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF07695.1"  
/db\_xref="GI:6382384"  
/translation="MEDIYMSKNNKKKFLNDRTEHNLIDNLSNLDTLNKLNVYKD  
LKEQNLNSLEIDTKIORMRILYKTRKLYKVDGKFSQYLKTFVAKTQAFYL  
KLYSKILGSGSITIKELGENIKKHIFKKKSPDFNNNKKSLDNKINISIRPNRDK  
ELYEFVKQDINKRASIIFKELFQNRKRDVLDTIVTYDNERN"  
15349..15813  
/gene="BBU07"  
CDS 15349..15813  
/gene="BBU07"

gene /note="similar to PID:1174341 PID:1174349 GB:AE001584  
percent identity: 81.71; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF07696.1"  
/db\_xref="GI:6382385"  
/translation="MIEILKTKRTEIKAKNNKHFTKSCSKEQKELKILCNTQKE  
LKSGNSQPLETNFOKIYENKYPHFIIENHKYSDLSYIKRLEKSIKIKENPQK  
DESLKINTFHFIEQLKKKEINTEILKPLVKEYLNQKKIKYKVFDTYTR"  
15791..16084  
/gene="BBU08"  
CDS 15791..16084  
/gene="BBU08"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF07702.1"  
/db\_xref="GI:6382391"  
/translation="MHITRGDNIKKYLSHTKVLTKDTFKSVFFRFLDKKNVIF  
DILKVAQLYKELINSEKTYQVVVEIENKELGYSINHRILIKVSSKSIYY"  
complement(16228..16548)  
/gene="BBU09"  
CDS complement(16228..16548)  
/gene="BBU09"  
/note="similar to GB:U03641 PID:458212 percent identity:  
66.22; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF07697.1"  
/db\_xref="GI:6382386"  
/translation="MGIFYQRECKTMSKAVDEIYCHSCGKTKKDAEICISGVRNK  
QTEYNKVVFLCLFLGLYGVHRFVVGKIGTGMLYLTFTGFLYVGLALIDIRIATNK  
FECK"  
16603..16800  
/gene="BBU10"  
CDS 16603..16800  
/gene="BBU10"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF07703.1"  
/db\_xref="GI:6382392"  
/translation="MVFLIEQLKKTNIETLPIREYLNQKKIKYKNTFGIYYLEL  
LEILKNRHDLNLEKSKVW"  
16886..17878  
/gene="BBU11"  
CDS 16886..17878  
/gene="BBU11"  
/note="similar to SP:Q10618 PID:132425 GB:AL123456  
percent identity: 53.61; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF07698.1"  
/db\_xref="GI:6382387"  
/translation="MIKTPFIISEDQFKIALKLMGEIIVFTDTTVYVAGIANYIDEYA  
VRMNLKKSINPEMLYDTEIKELSSHPKSAKILMEKFSQPLTYVLKKSIAK  
LPGFNNNDTVAIRPKNKTALNIRKLNPLVSSANLSKRPSSTFSMALKELNG  
LYGLIMPKKNDNIGIESTILEFDSKDNVILIRPCALITKKIEKVLGSRKYVYAK  
EKKDLKMFSTRLRIPVHFRGDNIRKLTNKLITRDTFKFTFFSFLWDKKNL  
ILFDSLIEYAENLYKELVNAEKYDQVIVEIENEELGAIYNNRIVKASNNFN"  
17918..18365  
/gene="BBU12"  
CDS 17918..18365  
/gene="BBU12"  
/note="this region contains an authentic frame shift and  
is not the result of a sequencing artifact; similar to

```

GB-L131615 GB-L131418 PID:520783 PID:551742 percent
identity: 61.90; identified by sequence similarity;
putative"

Query Match          5.3%; Score 42.6; DB 2; Length 18753;
Best Local Similarity 53.3%; Pred. No. 0.67;
Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 553 GTGCAGTCGCATTGTCTCTCTTTTCTTGATGGTTGGGAGCAGATCGCATTTTACCTTGGGA 612
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16392 GTGCTATTTTACTATGCTTATTTCTTGTTATTTAGGAGTTCACAGATTTTATGTAGGT 16333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 613 TACCTCGTTGGGTTGTTTAAAGTTTTCACCTGTAGGTTTGTGGAAATTTGGAGCCCTA 672
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16332 AAAATAGGAACCTGGTATGTTTATACCTATTTACATTTGGGATTTTATATGTTGGAGCTTTA 16273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 673 ATTGATTCATCTTATTTTCAATGCAGATTCTTGGACCTTCAGATGGA 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16272 ATCGATCTTATTAGAAATAGCAACAACAGTTTGAATGTAATAAGAAA 16224
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AB017069/c
LOCUS
DEFINITION
AB017069 88128 bp DNA PLN 20-NOV-1999
Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MNJ8,
complete sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui pl
clone:MNJ8.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (sites)
Nakamura.Y.
TITLE
Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
JOURNAL
Unpublished (1998)
REFERENCE
2 (bases 1 to 88128)
Nakamura.Y.
AUTHORS
Direct Submission
TITLE
Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamu@kazusa.or.jp, Tel:++81-438-52-3935,
Fax:++81-438-52-3934)
FEATURES
Location/Qualifiers
source
1..88128
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MNJ8"
/clone_lib="Mitsui pl"
BASE COUNT 29506 a 15101 c 14971 g 28550 t
ORIGIN

Query Match          5.1%; Score 41.6; DB 7; Length 88128;
Best Local Similarity 51.6%; Pred. No. 1.7;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 348 ACAATATTTGTAAAGATCCAAAATAATATGACGCTACGCAAGACCATTAAGTGTAC 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28303 ACAATACATATTATAAGCAAGATATGACGCACTAGTAAACAGATATCAAGATCTAATC 28244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 408 AAATACACAGCTCATGTTTCTCTGTTTCCAGCACCAACATACATCTGTAAGGATTCAG 467
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28243 AAACCATCTATTAACTAAGCAATTTTCCAGTACAAACAACTCTAAAAGGCATGATT 28184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 468 TGGCAATGAACACATTTTACTGGGAACAGAGTTGGTTTTTTTTCGAAGCCCATATCTTCCG 527

```





QY 638 TTTCAGCTAGGCTTTGTGGATTTGGAGCCTAATGATTTCATCTTCTTTTCAATCC 697  
 Db 9017 ATATATTCTGAGAAGCTCGAAGTGTGCACCTTTATATCATCTACTCATCTATTATTAT 8958  
 QY 698 AGATTGTTGGACCTTCAGATCGAAGTAGTCTTACATATATAGATTACTATGGAACACAGACTTA 757  
 Db 8957 GCCTGTTTATGCTTTAGTTCATCTATGCCAGTGATATGCTTTTAAACTTCTATTA 8898  
 QY 758 CAAGACTGAGTATTACATAAACAATTTAGAAAACCAATATATATCAATA 809  
 Db 8897 CTAAAGTAGTAAAGCAATAGATAAGGCTAAAGACATGCAATTCACA 8846

RESULT 14  
 BHV1CGEN 135301 bp DNA VRL 11-JAN-1999  
 LOCUS Bovine herpesvirus 1 complete genome.  
 DEFINITION AJO04801  
 ACCESSION AJO04801  
 VERSION AJO04801.1 GI:2653291  
 SOURCE complete genome.  
 ORGANISM Bovine herpesvirus type 1.1.  
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 Alphaherpesvirinae; Varicelloviruses.  
 REFERENCE 1 (bases 9702 to 103033)  
 AUTHORS Wirth,U.V., Fraefel,C., Vogt,B., Vlcek,C., Paces,V. and Schwytzer,M.  
 TITLE Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1  
 are 3' coterminal and encode a putative zinc finger transactivator protein  
 J. Virol. '66 (5), 2763-2772 (1992)  
 JOURNAL 92219360  
 MEDLINE 2 (bases 103034 to 111027; 127191 to 135300)  
 REFERENCE Schwytzer,M., Vlcek,C., Menekse,O., Fraefel,C. and Paces,V.  
 AUTHORS Promoter, spliced leader, and coding sequence for BICP4, the  
 TITLE largest of the immediate-early proteins of bovine herpesvirus 1  
 JOURNAL Virology 197 (1), 349-357 (1993)  
 MEDLINE 94025583  
 REFERENCE 3 (bases 111028 to 114234; 123984 to 127190)  
 AUTHORS Schwytzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.  
 TITLE BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA  
 which exhibits immediate early and late transcription kinetics  
 J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)  
 JOURNAL 94292919  
 MEDLINE 4 (bases 66900 to 96900)  
 REFERENCE Vlcek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D.,  
 AUTHORS Letchworth,G.J. and Schwytzer,M.  
 TITLE Nucleotide sequence analysis of a 30-kb region of the bovine  
 herpesvirus 1 genome which exhibits a colinear gene arrangement  
 with the UL21 to UL4 genes of herpes simplex virus  
 Virology 210 (1), 100-108 (1995)  
 JOURNAL 95313343  
 MEDLINE 5 (bases 1 to 31444)  
 REFERENCE Schwytzer,M., Sygder,D., Vogt,B., Lowery,D.E., Simard,C.,  
 AUTHORS Labouliere,S., Misra,V., Vlcek,C. and Paces,V.  
 TITLE Gene contents in a 31-kb segment at the left genome end of bovine  
 herpesvirus-1  
 Vet. Microbiol. 53 (1-2), 67-77 (1996)  
 JOURNAL 97164286  
 MEDLINE Accession# 254206  
 REFERENCE 6 (bases 30801 to 67800)  
 AUTHORS Schwytzer,M., Vlcek,C., Lowery,D.E., Bello,L.J., Meyer,G. and  
 Misra,V.  
 TITLE Gene contents in a 37-kb segment centered in the UL part of the  
 bovine herpesvirus 1 genome: the last gap  
 Unpublished  
 JOURNAL Accession# 278205  
 REFERENCE 7 (bases 96901 to 99695)  
 AUTHORS Letchworth,G.J. and Kutish,G.F.  
 TITLE DNA sequence of the BHV-1 UL1 to UL3.5 genes  
 Unpublished  
 JOURNAL 95313343  
 MEDLINE 8 (bases 114235 to 122983)  
 REFERENCE Goltz,M., Buhk,H.J., Brohl,H., Lewin,M., Mankertz,A., Boerner,B.,  
 AUTHORS Borchers,K. and Weigelt,W.

TITLE Nucleotide sequence of the HindIII O and K fragments located in the  
 US region of the bovine herpesvirus 1 genome  
 JOURNAL Unpublished  
 REMARK Accession# 298199  
 REFERENCE 9 (bases 121402 to 123983)  
 AUTHORS Schwytzer,M.  
 TITLE Glycoprotein E and US9 genes of BHV1  
 JOURNAL Unpublished  
 REFERENCE 10 (bases 1 to 135301)  
 AUTHORS Schwytzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J.,  
 Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.  
 TITLE Complete DNA sequence of bovine herpesvirus 1  
 Unpublished  
 REFERENCE 11 (bases 1 to 135301)  
 JOURNAL Direct Submission  
 TITLE Schwytzer,M.  
 JOURNAL Submitted (08-SEP-1997) Institute of Virology, Faculty of  
 Veterinary Medicine, University of Zurich, Winterthurerstrasse  
 266A, Zurich CH-8057, Switzerland  
 FEATURES  
 source  
 1. 135301  
 /organism="Bovine herpesvirus type 1.1"  
 /db\_xref="taxon:79889"  
 <1...472  
 /note="3' part of intron spliced out from transcript over  
 covalently joined genome ends, leading to immediate-early  
 RNA 1.5 and circ protein"  
 order(1. 4014,102918. 111027,127191. 135300)  
 /organism="Bovine herpesvirus type 1.2 (strain K22)"  
 /viiion="K22"  
 /strain="K22"  
 /db\_xref="taxon:31519"  
 235. 430  
 /rpt\_unit=235. 248  
 /rpt\_type=DIRECT  
 466. 1280  
 /note="transcribed from alternative late promoter"  
 /evidence=experimental  
 /product="LR1.1 (circ)"  
 486. 1229  
 /note="immediate-early and late myristylated protein  
 (virion)"  
 /codon\_start=1  
 /product="circ protein"  
 /protein\_id="CAA06078.1"  
 /db\_xref="GI:2653292"  
 /db\_xref="SPTREMBL:Q01342"  
 /translation="MGARASAPAGPPPAHVLDAISGTTDLPGDEAVFVSCPTT  
 RPVYHMRGRGTAHTTPVHFVGRAYAILPCRKFMLYLMRGAVYGYEPTTGLHRLADS  
 LHDFLTATAGLQRLDCLDVTVLDAQMDVFTTPEILIELEADPAFPPPPARARS  
 TLRRASMRPAPRTFCPHOLLAEGSLDLCSPQAAAPGCSLLPACDSGACPCDAGE  
 TARDCTADAARAPSPGALSRYSSVRSVFF"  
 1259. 1264  
 /gene="Circ"  
 /gene="Circ"  
 complement(1649. 1654)  
 /gene="UL52"  
 complement(1649. 7237)  
 /gene="UL52"  
 complement(1658. 2860)  
 /gene="UL54"  
 complement(1658. 2860)  
 /gene="UL54"  
 /note="early BICP27 affecting mRNA processing (cell  
 nucleus)"  
 /codon\_start=1  
 /product="UL54 protein"  
 /protein\_id="CAA06079.1"  
 /db\_xref="GI:2653293"  
 /db\_xref="SPTREMBL:Q01354"  
 /translation="MADPEIATLSTASEDDLSLFGSDREDEDEAPSLAPLRSVVGO  
 VRKRLEGADEPMPAEPGGAAGGPAEPARRARVRPRRRPRRRPAGE

```
ORSRGPAKREAAATSHGSGGGAARSGISLRLARSALAEAAQARATERTAVTAFAGA
RLDLMRPVONGFRAAGVSPWAAVLDFGAEOVFPEGRVVTWETLTFHGRDLRMEFVR
PHAAQALRLDLVRSANLVDALASADECLTWKFIATKNLRLKTKDPITVATAGAE
ENLKLAPFLRLGRGLSLLEELCAARLSLATCPASVFMVLMARLSRAVRSGAE
CVPLEVTGDPAPPEYIPGTCAVLIDALDTHKQACDSMTCKLVANFTLVVPVYHKG
YFYCNEIF"
/feature "complement(3040..4038)
/feature "gene="UL53"
/feature "complement(3040..4038)
/feature "gene="UL53"
/feature "codon_start=1
/feature "product="glycoprotein K"
/feature "protein_id="CAA06080.1"
/feature "db_xref="GI:2653294"
/feature "db_xref="SPTREMBL:O65816"
/feature "translation="MLGGRTVNLAAALHTAHLALALWALAAARCORCAVRATARN
GSLRWELSPGAVYWGANNATLAADPCRHAVVQHIPPGLLDGEALHGRVAVAG
ARDCAVYWCAGRGGLLAWLIVAFVLRQRRMFGICRNDADFSLPGGYTLNFAAA
ALAAVGHGPTVKLARLCELSARRALAVDFRLDPLGCANRPRAALPLAEGFARLG
ARIAAGSVGITHPCAAAYPLYLKIWVHVVALFAGLELVLLYKRRRGTCAAGD
GDGSGIRKVCNCCSTLLAGLLVKALYLAAVGVVIALHYEHLRLRLIGAQT"
/feature "complement(4013..7237)
/feature "gene="UL52"
/feature "codon_start=1
/feature "product="component of DNA helicase/primase complex"
/feature "protein_id="CAA06081.1"
/feature "db_xref="GI:2653295"
/feature "db_xref="SPTREMBL:O65817"
/feature "translation="MDGPALEALQILYATDGCAGISFLMLLTQEPGRGAYVVS
NMPGLAACGESNDVALADARPLVAFYLLGGPPGGGSAATPRLARLHVHQOR
AGRARARTAPROGTARARARPRGRHLCAARGHDDHALIVATEQLARTGTADF
ARYEQTGTVTSALRAGPAGRLASLIHHEKHTVAAVRYLSNGATPFWFLSKFG
PGEKTLVATRFYVFOAERAGDAATVDLQAVRDCLATYAAVAPNPSPGLAFDLYSFA
ALAAFCSSGARGAAGAPAYVAARIEADLAEVRLREYIDHRSLSKVADPEFA
YVLYAFGRNVQITEHLRATATEPPDGGALAPLAGVSLRERAVDAFFRVRA
QFNQVSYLEQNVAVVRLPATAEAYARTAYARLLAANTANGPRTICDGAALRAA
LDHLEGAARCGWLHPAGPAGSAPVPGGLDAAALRVGAPDGGSPDTPACCGYSKR
LLELAAARADGAHDLGLFGARGTAAPTVYRVELPRGHOAFVADGDWAAVTSVA
ALDPDEVAVEVAAAARAEGRDALVERDARLTALLTERACAPAGRSRDLSPGAP
RDQYVNNRLENFARLATNIVLDVFLKRLPRGLDHLGAMRSFRGALALALAFP
ETIADAPHAFCFFYKSCPSAASNLGDGASPSWAPDRDAGWDEYESALARGEA
QYSCSDVDGAPAEETGGAGSAAALTACDDKMGFRVAVPAPYPIIAGGP
TLKGTARLVQAHMERLEPFAEMSYLRDFAFVDAVYTHGHSRLPFGKPDASNV
RWLLPFVYPERCDDVAAVAAHRDPHFHFAAPPAGGRRARHVGVEYVQFE
KRAAVNRAFGVRLVRLAAACVGDGDSGAEFADVIEALAYMAHYEQAA
REYQAGARLVNAKADWLLAQVLPARAGTAGFGCLRAGHGAQNAKRSVLSVDS
HRLCASLIQOCFAFKCSGNRLGTVFTVDSLGRAARYAARGAD"
/feature "order(4015..9664,15010..22094,26636..102917,
114235..123983,135301)
/feature "organism="Bovine herpesvirus type 1.1 (strain Cooper)"
/feature "villon
/strain="Cooper"
/feature "db_xref="taxon:10323"
/feature "7236..7981"
/feature "gene="UL51"
/feature "7236..7967"
/feature "gene="UL51"
/feature "codon_start=1
/feature "product="villon protein"
/feature "protein_id="CAA06082.1"
/feature "db_xref="GI:2653296"
/feature "db_xref="SPTREMBL:O65818"
/feature "translation="MALORLVNLCGRGGTRDAPQYEPPLARRLSGPTLFRLOEAVVA
VSALLPALTVEDVADSGTRRLAKASLARTYICQNIETCSKHQAACSDAIVTA
VTRHIQDAQRMDCTDLAALLQMYHSVAGVETGTDMSYDQAIRMAAESINVDYAVL
ERALGIOAQAGAEAAASPTAVGAEARAPPLPQPAASAPVAVSPAPPVPAEAA
ATKPPSRAGRAAAASSPAVLQAA"
/feature "7976..7981"
/feature "polyA_signal
/feature "gene="UL51"

Query Match
Best Local Similarity 48.1%; Score 39.8; DB 16; Length 135301;
Matches 113; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
```

```
QY 76 GGCTCTATCTATGCGAGAGTCCCTTTAAGAACTCGCCCTGTGTCCTCCCTCCG 135
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106797 GCGCGGCGCGTCCGCGCGCGCGGTAGCACAGCGGCGCGCGGTTCGCGCCG 106856

QY 136 CTCTGGCGGAGCGGGAAGTGGCGAGAAAGTGTCCGCTCCCAAGATGCGGCC 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106857 CGCTGCGCGCGCGGCGGCGCGCGCTCTCCGCGCGCGAGGCGCGCGGCC 106916

QY 196 GCCTGCGCGTCTGCTGCTCGCGAGGCGCGTACGCGGCGAGCTCGTGGTGTCTG 255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106917 GCGCGCGCGCGCGGCGCGCGCGCTCGCGCGCGCGCTCGCGCGCGCTTGTGTTCTG 106976

QY 256 TGTTCGTCTCAGTCACTACAGACCCCTGGGGGCTGTTCACCTCCGCGCGGG 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106977 GTCCGCGCGCGTCCGCGCTCAGCCCTCGGCTCGGCTCGGCTCGGCTCGGCTCGS 107031

RESULT 15
BHVICGEN/LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BHVICGEN 135301 bp DNA VRL 11-JAN-1999
Bovine herpesvirus 1 complete genome.
AJ004801
AJ004801.1 GI:2653291
complete genome.
Bovine herpesvirus type 1.1.
Bovine herpesvirus type 1.1
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphanherpesvirinae; Varicellovirus.
Wirth,U.V., Fraefel,C., Vogt,B., Vilek,C., Paces,V. and Schwyzer,M.
Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
are 3' coterminal and encode a putative zinc finger transactivator
protein
J. Virol. 66 (5), 2763-2772 (1992)
JOURNAL
MEDLINE
92219360
REFERENCE
2 (bases 103034 to 111027; 127191 to 135300)
Schwyzer,M., Vilek,C., Menekse,O., Fraefel,C. and Paces,V.
Promoter, spliced leader, and coding sequence for BICP4, the
largest of the immediate-early proteins of bovine herpesvirus 1
Virology 197 (1), 349-357 (1993)
JOURNAL
MEDLINE
94025583
REFERENCE
3 (bases 111028 to 114234; 123984 to 127190)
Schwyzer,M., Wirth,U.V., Vogt,B. and Fraefel,C.
BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA
which exhibits immediate early and late transcription kinetics
J. Gen. Virol. 75 (Pt 7), 1703-1711 (1994)
JOURNAL
MEDLINE
94292919
REFERENCE
4 (bases 66900 to 96900)
Vilek,C., Benes,V., Lu,Z., Kutish,G.F., Paces,V., Rock,D.,
Letchworth,G.J. and Schwyzer,M.
Nucleotide sequence analysis of a 30-kb region of the bovine
herpesvirus 1 genome which exhibits a colinear gene arrangement
with the UL21 to UL4 genes of herpes simplex virus
Virology 210 (1), 100-108 (1995)
JOURNAL
MEDLINE
95313343
REFERENCE
5 (bases 1 to 31444)
Schwyzer,M., Styger,D., Vogt,B., Lowery,D.E., Simard,C.,
LaBoissiere,S., Misra,V., Vilek,C. and Paces,V.
Gene contents in a 31-kb segment at the left genome end of bovine
herpesvirus-1
Vet. Microbiol. 53 (1-2), 67-77 (1996)
JOURNAL
MEDLINE
97164286
REFERENCE
6 (bases 30801 to 67800)
Schwyzer,M., Vilek,C., Lowery,D.E., Bello,L.J., Meyer,G. and
Misra,V.
Gene contents in a 37-kb segment centered in the UL part of the
bovine herpesvirus 1 genome: the last gap
Unpublished
Accession# 278205
7 (bases 96901 to 99695)
Letchworth,G.J. and Kutish,G.F.
```

```

TITLE      DNA sequence of the BHV-1 UL1 to UL3.5 genes
JOURNAL    Unpublished
REFERENCE  8 (bases 114235 to 122983)
AUTHORS    Goltz,M., Buhk,H.J., Brolli,H., Lewin,M., Mankertz,A., Boerner,B.,
            Borchers,K. and Weigelt,W.
TITLE      Nucleotide sequence of the HindIII O and K fragments located in the
            US region of the bovine herpesvirus 1 genome
JOURNAL    Unpublished
REMARK     Accession# Z98199
REFERENCE  9 (bases 121402 to 123983)
AUTHORS    Schwytzer,M.
TITLE      Glycoprotein E and US9 genes of BHV1
JOURNAL    Unpublished
REFERENCE  10 (bases 1 to 135301)
AUTHORS    Schwytzer,M., Paces,V., Letchworth,G.J., Misra,V., Buhk,H.J.,
            Lowery,D.E., Simard,C., Bello,L.J., Thiry,E. and Vlcek,C.
TITLE      Complete DNA sequence of bovine herpesvirus 1
JOURNAL    Unpublished
REFERENCE  11 (bases 1 to 135301)
AUTHORS    Schwytzer,M.
TITLE      Direct Submission
JOURNAL    Submitted (08-SEP-1997) Institute of Virology, Faculty of
            Veterinary Medicine, University of Zurich, Winterthurerstrasse
            266A, Zurich CH-8057, Switzerland
FEATURES   Location/Qualifiers
            source
              1..135301
              /organism="Bovine herpesvirus type 1.1"
              /db_xref="taxon:79889"
              <1..472
            intron
              /note="3'part of intron spliced out from transcript over
              covalently joined genome ends, leading to immediate-early
              RNA 1.5 and circ protein"
              order(1..4014,102918..111027,127191..135300)
            source
              /organism="Bovine herpesvirus type 1.2 (strain K22)"
              /virus
              /strain="K22"
              /db_xref="taxon:31519"
              235..430
            repeat_region
              /rpt_unit=235..248
              /rpt_type=DIRECT
              456..1280
            mRNA
              /note="transcribed from alternative late promoter"
              /evidence=experimental
              /product="LRI.1 (circ)"
              486..1229
            CDS
              /note="immediate-early and late myristylated protein
              (viriion)"
              /codon_start=1
              /product="circ protein"
              /protein_id="CAA06078.1"
              /db_xref="GI:2653292"
              /db_xref="SPTREMBL:001342"
              /translation="MGRASAPAGPPPAHVALLDALSGGTIDLPDGDEAVFVSCPTT
              RPVHHMRRTAHTTPVHVGRAIILPCRKFMLYLMRGAVYGEFTTGLHRLADS
              LHDLFTTAGQQQLDHLCDVLDQMDPVTFTTEILLIEADPAFPPPPSARARS
              TLRRASMRPRTFCPHQLAEGSILDLCSPEQAAPGCSILLPACDSGDAACPCDAGE
              TARDCTADAAAPSPGALSRYSSVSRYFF"
              1259..1264
              /gene="circ"
              1259..1264
            polyA_signal
              /gene="circ"
              complement(1649..1654)
              /gene="UL52"
              complement(1649..7237)
              /gene="UL52"
              complement(1658..2860)
              /gene="UL54"
              complement(1658..2860)
              /gene="UL54"
              /note="early BICP27 affecting mRNA processing (cell
              nucleus)"
              /codon_start=1
              /product="UL54 protein"
              7976..7981
            polyA_signal
              7976..7981
            gene
              /strain="Cooper"
              /db_xref="taxon:10323"
              7236..7981
            polyA_signal
              /gene="UL59"
              7236..7967
            gene
              /gene="UL51"
              /codon_start=1
              /product="viriion protein"
              /protein_id="CAA06082.1"
              /db_xref="GI:2653296"
              /db_xref="SPTREMBL:065818"
              /translation="MALQRLVGLCGRGGTRDAPQVEPLARRLSGPTFLRLOEAVVA
              VSALLPAPLVEDVARSDGTRRLAKAQSGLARTYICORNIECLSKHOACSDASITA
              VYTKHQDAQMRDTCCLAQLQWYHSVGADGTTTDMQDQAIRMAAESNIYADYAVL
              ERALGIAQAGAGAAARASPTAVGAEARAPPLPQPAASAPVAVSPAPPVAPAEA
              ATKPPSRAGRAAAAASSPAVLQIAA"
              7976..7981
            polyA_signal
              7976..7981
            gene
              /protein_id="CAA06079.1"
              /db_xref="GI:2653293"
              /translation="MADPEIATLSTASESDSLSGSDEEDEAPSLAPALRSVVGQ
              VRKKEGADEPMPAEPPEGGAGSGDGAEPARRARVRPRPRPRRPPRQPAGE
              RSRGPKRAKREALATSSHGGGAARSIGSLRLASLAEOARATSERVTA'AVFAGA
              RLDLMRPVONGGFRAGVSPAAVLDFAEQFVPEGRVWTETLHFGRDLYRMEVVR
              PHAAARALRDLYLRSANLVDALADECTWCKFIATKNLRLTKPIVATAGAVL
              ENLRKALPFLRCYLRGLSPLEELCAARLSLATCPASFMVMLARLSRAVRSGAE
              CVPLEIVTGDAPEEYIPGTCVAGLIDALDTHKQACDSMTCKLVANETLVPVYMHKG
              YFYCNEIF"
              complement(3040..4038)
              /gene="UL53"
              complement(3040..4038)
              /gene="UL53"
              /codon_start=1
              /product="glycoprotein K"
              /protein_id="CAA06080.1"
              /db_xref="GI:2653294"
              /db_xref="SPTREMBL:Q65816"
              /translation="MLLGGRTVNLAALTAHLALALWALAARCAQCACVRATARN
              GSLRELRSPCAVYVWGANNATLAADAPCHAVVOHIPPGLLDGDEALHGRVRAVG
              ARDCRAVLCQAQARGGLLAWLYVAVYLRQERMGCLCRNDADFLSGGYFLNFAAA
              ALAAVGHGPTTKLARLMCELSARRALAVDFRLDPCAMRPRAALLGAEFARLG
              ARIAAGSVGITHPCAAAPPLYLKIWAVHVALFAGLELVSLLYKPKRRRGCTGADG
              GDGSGIRKYCVNCCSTLLAGLLVKA'LYLAIVGGVITALLHYEHLRLLLGAQT"
              complement(4013..7237)
              /gene="UL52"
              /codon_start=1
              /product="component of DNA helicase/primase complex"
              /protein_id="CAA06081.1"
              /db_xref="GI:2653295"
              /db_xref="SPTREMBL:Q65817"
              /translation="MDGPALEDALQILYATDGAIGFSLMLLTGOEPRGCAYVVS
              NPGRLAACLGESPNVVALADARPLVAYLLGGPGGSAAT'PRLARLVHLQQR
              AGRARARRTA'PROGTPTRARAARPRGRHILCAAGHALLVATEQIAPRTGRTADD
              ARYVEGQTTVRSALRPAAGRLASLYIHHEHTVAYRRLYSNGATPFWFLSKFG
              PGKTLVLA'FRYVFAERAGDAATYDLOAVRDCIATYAAVAPNPSGLAFDLYSFA
              ALAAFCRSGVARGAAGVAAVAA'RTADLA'EVRCUREYIDHRRSLKVNADPEFVA
              YVLYAFEGFNERQIT'ELHRAVTA'ETPDGGEALAPRLAGVSLREARVADAF'FRVRA
              QFNQSYIEQNVA'VVRPLPA'AEATARATYAKLAA'ANGPRTICDGAALURAA
              LDHLEGAARFEGVHLHAPGASAVPGDLAALRVGAPDGGSPDTPACCGVSKR
              LLELAAAARDGAHP'LDLFGARGAATPYVYV'ELPRGHAFAVADGDWAAVTSVA
              ALDPDEVA'VEVAAAAA'ESRDALVERDARLTALLTERAACAGRSRSLGPAAP
              RDQYVNRNELFARLAVTNIVLDVDFRLKPLPGDLHGA'RSFRCALALALFEP
              EIAADAMP'HCFFYKSNACPSAAASNLGCDASPSWAPDRDAGWDEYESALARGEA
              QICEDVSDPAGARPEETGGAGGASAAALTACGDDKMGFRVAVPYVPYPIITAGGP
              TLKGIALVQHVMLEP'FAEAMSYLRDFAVDAAYTHGHSRLR'LPFGKPDASNY
              GRWLLPFYVVP'ERCDDVAA'FAAHRDPHFHFAAPAGGARRARRAHVGEYVSFFE
              RKAAVNRAA'FHGVRSLLRAA'VGVGDGDRGAE'EFADYIEALAPYMAEHYQAAA
              REYOGAGARLVA'AKDMLAOLV'PARAGTA'OGFCGLRAAHGA'QONKARSLVSLVSDS
              HDRCASLIQOCFATKCGSNRLGTVFTV'VDSGRAARV'AAAGADC"
              order(4015..9664,15010..22094,28636..102917,
              114235..123983,135301)
              /organism="Bovine herpesvirus type 1.1 (strain Cooper)"
              /virus
            polyA_signal
              /strain="Cooper"
              /db_xref="taxon:10323"
              7236..7981
            gene
              /gene="UL59"
              7236..7967
            polyA_signal
              /gene="UL51"
              /codon_start=1
              /product="viriion protein"
              /protein_id="CAA06082.1"
              /db_xref="GI:2653296"
              /db_xref="SPTREMBL:065818"
              /translation="MALQRLVGLCGRGGTRDAPQVEPLARRLSGPTFLRLOEAVVA
              VSALLPAPLVEDVARSDGTRRLAKAQSGLARTYICORNIECLSKHOACSDASITA
              VYTKHQDAQMRDTCCLAQLQWYHSVGADGTTTDMQDQAIRMAAESNIYADYAVL
              ERALGIAQAGAGAAARASPTAVGAEARAPPLPQPAASAPVAVSPAPPVAPAEA
              ATKPPSRAGRAAAAASSPAVLQIAA"
              7976..7981
            polyA_signal
              7976..7981
```

```
/gene="UL51"
```

Query Match	4.9%	Score 39.8	DB 16	Length 135301	
Best Local Similarity	48.1%	Pred. No. 5.5			
Matches 113	Conservative	0	Mismatches 122	Indels 0	Gaps 0
QY	76	GGCCTCTATCCTATGCAGGTGCCCTTTAAAGAACTCGCCCTGTGGCCCTTCTCCCTCCCG	135		
Db	131421	GGCGGCGCCCGTCGCGCGCGCGCGGTAGCCAGCGGGCGCGGTTCGCGCGCG	131362		
QY	136	CTCCTGGGGCGAGCGGAAGTGGCGAGAAATGTTCGGTCTCCAAAGATGGCGGCC	195		
Db	131361	CGTGTGGCGCGCGCGGGGGGGGCGCGTCTCCGGCGCGAGAGCGCGCGGGGCGCGGC	131302		
QY	196	GCCGTGGCGCTGTGTCCTGTCTGTCGGAGGCCGTGACGGCCACACACTGTGTGGTCTCGT	255		
Db	131301	GGCGGGCCCCGGGGCGCGGGCCCCGCCCTTCGCGGGGCGCTTTCGTGTCGT	131242		
QY	256	TGFTTCGTCTACGTACTACAGGACCTGGGGGCGCTGTTCACCTCCGCGGGG	310		
Db	131241	GTCCCGGCCCAAGTCCCGGGCTCAGCCTCGGCCCTTCGGCCTCAGCGTGGCGCTCGG	131187		

Search completed: February 11, 2000, 01:45:19  
Job time: 3932 sec



|||||  
Db 241 SSIIDYGTSLRLSITNETFRKTQLYP 269  
|||||  
RESULT 2  
Y12358  
ID Y12358 standard; Protein; 139 AA.  
AC Y12358;  
DT 17-JUN-1999 (first entry)  
DE Human 5' EST secreted protein SEQ ID NO:389.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
OS Homo sapiens.  
WO9908548-A2.  
11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153778/13.  
DR N-PSDB; X41191.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 27; Page 714-715; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
Sequence 139 AA;  
Query Match 49.4%; Score 133; DB 1; Length 139;  
Best Local Similarity 100.0%; Pred. No. 9.1e-115;  
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 69 SGPSAPEAVTARLVGLWVSVTTGPMGAVATSAGGESLKCEDLKVGQYICKDPKINDA 128  
Db 7 SGPSAPEAVTARLVGLWVSVTTGPMGAVATSAGGESLKCEDLKVGQYICKDPKINDA 66  
Qy 129 TQEPVNCNTYTAHVSCFPAPNITCKDSSGNETHTGTGNEVGFKPISCRNVNGYSYKVA 188  
Db 67 TQEPVNCNTYTAHVSCFPAPNITCKDSSGNETHTGTGNEVGFKPISCRNVNGYSYKVA 126  
Qy 189 LSLFLGWLGDREF 201  
Db 127 LSLFLGWLGDREF 139  
RESULT 3  
Y12426  
ID Y12426 standard; Protein; 148 AA.  
AC Y12426;  
DT 17-JUN-1999 (first entry)  
DE Human 5' EST secreted protein SEQ ID NO:457.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;

forensic; gene therapy; chromosome mapping; signal peptide;  
upstream regulatory sequence; cytokine activity; cell proliferation;  
differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
OS Homo sapiens.  
WO9908548-A2.  
11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153778/13.  
DR N-PSDB; X41259.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 27; Page 763-764; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
Sequence 148 AA;  
Query Match 30.5%; Score 82; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 6.2e-68;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 73 APEAVTARLVGLWVSVTTGPMGAVATSAGGESLKCEDLKVGQYICKDPKINDATQEP 132  
Db 11 APEAVTARLVGLWVSVTTGPMGAVATSAGGESLKCEDLKVGQYICKDPKINDATQEP 70  
Qy 133 VNCNTYTAHVSCFPAPNITCKD 154  
Db 71 VNCNTYTAHVSCFPAPNITCKD 92

RESULT 4  
W1297  
ID W1297 standard; Protein; 457 AA.  
AC W1297;  
DT 02-FEB-1999 (first entry)  
DE Lucilia cuprina ecdysteroid receptor.  
KW Ecdysteroid receptor; insecticide; biological control;  
KW gene therapy; Australian blowfly.  
OS Lucilia cuprina.  
FT Key Location/Qualifiers  
FT Domain 1..300  
FT /label= A/B  
FT Domain 301..366  
FT /label= C  
FT /note= "DNA binding domain"  
FT Domain 367..453  
FT /label= D  
FT Domain 545..674  
FT /label= E  
FT /note= "hormone binding domain"  
FT Domain 675..757  
FT /label= F

FT Peptide 379...385  
 FT /label= NLS  
 FT /note= "nuclear localisation signal"  
 FT 387...391  
 FT /label= NLS  
 FT /note= "nuclear localisation signal"  
 FT 495...536  
 FT /note= "helix-turn-zipper motif"  
 PN W09835550-A2.  
 PD 20-AUG-1998.  
 PF 16-FEB-1998; NZ0018.  
 PR 16-FEB-1997; NZ-314239.  
 PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 PA (NZWO-) NEW ZEALAND WOOL BOARD.  
 PI Heath ACG, Huang SQ;  
 DR WPI: 98-520754/44.  
 DR N-PSDB; V60302.  
 PT New Lucilia cuprina ecdysteroid receptor, its fragments and related  
 PT nucleic acid - vectors, transformed cells, ligands and antisense  
 PT systems and for recombinant protein production  
 PT Example 2; Fig 4; 50pp; English.  
 CC This is the amino acid sequence of novel claimed Lucilia cuprina  
 CC ecdysteroid receptor (EcR), as deduced from the sequence of an  
 CC isolated polynucleotide (see V60302). Also claimed are peptides  
 CC containing one or more of the A/B, C, D, E or F (especially C or E)  
 CC domains of the ecdysteroid receptor, nucleic acid encoding them,  
 CC expression vectors, an inducible expression system, ligands able to  
 CC bind the receptor or its peptides, and a vector encoding antisense  
 CC RNA. Vectors that express the receptor, or its C or E domains,  
 CC vectors containing antisense sequences, and vectors in which a gene  
 CC encoding the ecdysteroid receptor is inactivated are used as  
 CC insecticides, particularly against flies and specifically L. cuprina  
 CC (Australian blowfly, a vector of human and animal diseases).  
 CC Overexpressing the receptor induces premature moulting while  
 CC antisense RNA prevents moulting. Ligands can also be used as  
 CC insecticides and vectors are particularly used in baits for adult  
 CC flies. Also contemplated is germline transformation of flies using  
 CC a transposon encoding the receptor or its C or E domains. Host  
 CC cells are used to produce recombinant receptor, and inducible  
 CC expression systems are used for steroid-responsive expression of  
 CC other genes, e.g. in gene therapy (where expression will be induced  
 CC by administration of insect steroids) or for large scale protein  
 CC production.  
 SQ Sequence 757 AA;

Query Match 3.0%; Score 8; DB 1; Length 757;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 LGGGGSGS 54  
 Db 147 LGGGGSGS 154

RESULT 5  
 R03623  
 ID R03623 standard; protein; 283 AA.  
 AC R03623;  
 DT 07-AUG-1990 (first entry)  
 DE Zucchini yellow mosaic virus (ZYMV) coat protein.  
 KW ZYMV; curcubitaceae; solanaceae; leguminosae; caricaceae;  
 KW potyvirus; ds..  
 OS Papaya ringspot virus.  
 PN W09002184-A.  
 PD 8-MAR-1990.  
 PF 20-JUL-1989; 003094.  
 PR 19-AUG-1988; US-234412.  
 PR 14-MAR-1989; US-323536.  
 PR 19-JUN-1989; US-368710.  
 PA (CORR) Cornell Res Found Inc, (UPJO) The Upjohn Company.  
 PI: Quemada H, Slightom JL, Gonsalves D, L'Hostis B;

DR WPI: 90-099408/13.  
 DR N-PSDB; Q03671.  
 PT Potyvirus coat protein genes - used to produce transformed plants  
 PT resistant to viral infection by potyvirus and related viruses.  
 PS disclosure; p; English.  
 CC Plant cells may be transformed with coat protein to render them resistant  
 CC to attack by potyvirus and related viruses from whence the sequence is  
 CC derived - papaya ringspot virus, Watermelon mosaic virus II and Zucchini  
 CC yellow mosaic virus.  
 CC The promoter and polyadenylation signals are from the Cauliflower mosaic  
 CC virus 35s promoter, and initiation region is derived from the Cucurbit  
 CC mosaic virus coat protein or RUBISCO gene.  
 SQ Sequence 283 AA;

Query Match 2.6%; Score 7; DB 1; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 GSGSGGK 57  
 Db 32 GSGSGGK 38

RESULT 6  
 R07661  
 ID R07661 standard; protein; 828 AA.  
 AC R07661;  
 DT 25-FEB-1991 (first entry)  
 DE Incomplete form of human RF-X protein.  
 KW MHC class II; down regulation; autoimmune disease.  
 OS Homo sapiens.  
 PN W09012812-A.  
 PD 01-NOV-1990.  
 PF 18-APR-1990; E00625.  
 PR 18-APR-1989; EP-106944.  
 PR 14-AUG-1989; EP-115008.  
 PA (MACH/) Mach B.  
 PI Mach B;  
 DR WPI: 90-348429/46.  
 DR N-PSDB; Q06459.  
 PT Purified proteins and compns. - regulate expression of MHC class II  
 PT genes and bind to controlling DNA sequences.  
 PS Claim 6; Page 40; 83pp; English.  
 CC The sequence encodes a protein which regulates the expression of  
 CC MHC class II genes by binding to DNA sequences which control this  
 CC expression. The protein causes down-regulation, useful for the  
 CC prevention and treatment of autoimmune diseases such as Insulin  
 CC Dependent diabetes, Multiple Sclerosis, Lupus Erythematosus and  
 CC Rheumatoid Arthritis. The protein can also be used for screening  
 CC and identifying substances capable of inhibiting the expression  
 CC of the MHC II genes.  
 CC See also R08338, R08390 and R08391.  
 SQ Sequence 828 AA;

Query Match 2.6%; Score 7; DB 1; Length 828;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGS 54  
 Db 245 GGGGSGS 251

RESULT 7  
 R08338  
 ID R08338 standard; protein; 979 AA.  
 AC R08338;  
 DT 25-FEB-1991 (first entry)  
 DE Complete form of human RF-X protein.  
 KW MHC class II; down regulation; autoimmune disease.  
 OS Homo sapiens.

PN WO9012812-A.  
 PD 01-NOV-1990.  
 PE 18-APR-1990; E00625.  
 PR 18-APR-1989; EP-106944.  
 PR 14-AUG-1989; EP-115008.  
 PA (MACH/) Mach B.  
 PI Mach B;  
 DR WPI; 90-348429/46.  
 DR N-PSDB; Q06470.  
 PT Purified proteins and compns. - regulate expression of MHC class II genes and bind to controlling DNA sequences.  
 PS Claim 6; Fig 10; 83pp; English.  
 CC The protein which regulates the expression of MHC class II genes by binding to DNA sequences which control this expression. It causes down-regulation, useful for the prevention and treatment of autoimmune diseases such as Insulin Dependent diabetes, Multiple Sclerosis, Lupus Erythematosus and Rheumatoid Arthritis. It can also be used for screening and identifying substances capable of inhibiting the expression of the MHC II genes. It has a calculated mol. wt of 104727.  
 CC See also R07661, R08390 and R08391.  
 SQ Sequence 979 AA;

Query Match 2.6%; Score 7; DB 1; Length 979;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGSGS 54  
 DB 396 GGGSGS 402

RESULT 8  
 ID R08390 standard; protein: 437 AA.  
 AC R08390;  
 DT 25-FEB-1991 (first entry)  
 DE Truncated form of human FR-X protein.  
 KW MHC class II; down regulation; autoimmune disease; HLA promoter.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT domain 255..376  
 FT /label=DNA binding domain  
 PN WO9012812-A.  
 PD 01-NOV-1990.  
 PE 18-APR-1990; E00625.  
 PR 18-APR-1989; EP-106944.  
 PR 14-AUG-1989; EP-115008.  
 PA (MACH/) Mach B.  
 PI Mach B;  
 DR WPI; 90-348429/46.  
 PT Purified proteins and compns. - regulate expression of MHC class II genes and bind to controlling DNA sequences.  
 PS Claim 6; Fig 10; 83pp; English.  
 CC The protein which regulates the expression of MHC class II genes by binding to DNA sequences which control this expression. It causes down-regulation, useful for the prevention and treatment of autoimmune diseases such as Insulin Dependent diabetes, Multiple Sclerosis, Lupus Erythematosus and Rheumatoid Arthritis. It can also be used for screening and identifying substances capable of inhibiting the expression of the MHC II genes. A polypeptide comprising only the DNA binding domain (see feature table) will bind to the X box of the HLA class II promoter.  
 CC See also R08338, R07661 and R08391.  
 SQ Sequence 437 AA;

Query Match 2.6%; Score 7; DB 1; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGSGS 54

Db 245 GGGSGS 251  
 RESULT 9  
 ID R14702 standard; Protein: 236 AA.  
 AC R14702;  
 DT 01-FEB-1992 (first entry)  
 DE Vh-Lab-VI construction (9), single chain antibody.  
 KW SCA; SCAPA; trombus; t-PA; urokinase.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT region 1..113  
 FT /label= Vh  
 FT /note= "from MAB MA-15C5"  
 FT 114..128  
 FT /label= Lab  
 FT 129..236  
 FT /label= VI  
 FT /note= "from MAB MA-15C5"  
 FT 108..111  
 FT /label= Vh\_anchor\_region  
 FT 118..122  
 FT /label= VI\_anchor\_region  
 PN WO9116353-A.  
 PD 31-OCT-1991.  
 PE 21-APR-1991; E00767.  
 PR 23-APR-1990; EP-401090.  
 PA (CORV-) CORVAS INT NV.  
 PI Larocche Y, Holvoet P, Demaeyer M;  
 DR WPI; 91-339763/46.  
 PT New single-chain Mabs specific to esp. fibrin - used for imaging, and when connected to plasminogen activating portion used for lysis of thrombi.  
 PS Disclosure; Fig 6; 78pp; English.  
 CC The SCA can be labelled and used for the imaging of thrombi.  
 CC The SCA may be connected, via a 2nd linker (Lcd) to a plasminogen activating portion (PA-portion), pref. a catalytic domain of t-PA or urokinase, esp. scuPA, to form a thrombolytic agent (SCAPA).  
 CC The SCAPA has an increased half-life and has a lower mol.wt. than similar agents and is therefore expected to have a reduced immunogenicity and improved thrombus penetration.  
 CC See also Q14468-70 and R14694-703 and R15690-93.  
 SQ Sequence 236 AA;

Query Match 2.6%; Score 7; DB 1; Length 236;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGS 55  
 DB 114 GGGSGS 120

RESULT 10  
 ID R20227 standard; Protein: 223 AA.  
 AC R20227;  
 DT 22-APR-1992 (first entry)  
 DE PY53.3 (NCIMB 40308)  
 KW TDF gene; Y-chromosome; mt-box; DNA-binding protein;  
 KW sex determination; mating type; testis determining factor.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT binding\_site 77..156  
 FT /label= mt-protein  
 PN WO9200375-A.  
 PD 09-JAN-1992.  
 PF 28-JUN-1991; G01057.  
 PR 28-JUN-1990; GB-014446.  
 PR 13-JUL-1990; GB-015488.



```

PR 09-MAY-1991; GB-010085.
PA (IMCR ) IMP CANCER RES TECH.
PI Goodfellow PN, Lovell-Badge R;
DR WPI; 92-041557/05.
DR N-PSDB; Q20684.
PT Nucleic acid pY53.3, its fragments and oligo-nucleotide(s) - is
PT used for determining or controlling sex of e.g. embryos by
PT detecting Y chromosomes
PS Claim 24; Page 125 and Fig 19(1-3); 183pp; English.
CC The TDF gene comprising a Y-chromosome specific sequence contg.
CC the mating type box (mt-box). The mt-box encodes a functional testis
CC determining factor (mt-protein, see features) which when expressed at
CC the appropriate stage of embryo development results in testis
CC formation and subsequent growth of the embryo as male.
CC The sequence of the mt-protein is similar to the DNA-binding motif
CC of known DNA-binding proteins. There are a number of residues
CC conserved between human, rabbit and mouse mt-proteins, but not
CC conserved in DNA-binding proteins not associated with sex
CC determination at least at the stage of testis formation.
CC See also Q20680-81, Q20684-87, Q22530-31, Q22535-36 and Q22560-67.
CC Sequence 223 AA;

Query Match 2.6%; Score 7; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 187 VALSLFL 193
Db 12 VALSLFL 18

RESULT 11
R24048
ID R24048 standard; Protein; 461 AA.
AC R24048;
DT 08-DEC-1992 (first entry)
DE Heavy chain of 4D5 Fab IgG antibody - recognises HER-2 receptor.
KW humanised IgG antibody; human growth hormone; hGH; selection;
KW Chinaeric Homo sapiens.
OS WO209690-A.
PD 11-JUN-1992.
PF 03-DEC-1991; U09133.
PR 03-DEC-1990; US-621667.
PR 10-APR-1991; US-683400.
PR 14-JUN-1991; US-715300.
PR 08-AUG-1991; US-743614.
( GETH ) GENENTECH INC.
Bass S, Garrard LJ, Greene R, Henner DJ, Lowman HB;
Matthews DJ, Wells JA;
WPI; 92-217069/26.
DR N-PSDB; Q25592.
PT Selecting and enriching variant proteins - comprises fusing gene
PT encoding e.g. growth hormone to part of M13 phage coat protein
PT and mutagenising fusion prior to selection
PS Claim 46; Fig 11; 102pp; English.
CC This sequence represents the heavy chain Fab portion of the 4D5
CC humanised IgG antibody that recognises the HER-2 receptor. The 4D5
CC gene was inserted into the plasmid pSOL32, which had the DNA encoding
CC human growth hormone excised from it. The plasmid was used to transform
CC E. coli SR101. - See Q25592.
CC Recombinant antibody was then used in a RIA for the HER-2 antigen
CC (ECD).
SQ Sequence 461 AA;

Query Match 2.6%; Score 7; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 GGGSGSG 55
Db 1111111

```

```

Db 305 GGGSGSG 311

RESULT 12
R27097
ID R27097 standard; Protein; 484 AA.
AC R27097;
DT 02-MAR-1993 (first entry)
DE Translation of the signal:BPTI:mature III construct.
KW Bovine pancreatic trypsin inhibitor; M13; major coat protein;
KW tripartite gene; surface; gene III.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1..18
FT region 19..76
FT protein 77..456
FT protein 457..479
FT protein 484..484
/note= "signal peptide"
/note= "BPTI insertion"
/note= "mature gene III protein"
/note= "uncharged anchor region"
PN WO9215679-A.
PD 17-SEP-1992.
PF 28-FEB-1992; U01539.
PR 01-MAR-1991; US-664989.
PR 17-JUN-1991; US-715834.
PA (PROT-) PROTEIN:ENG CORP.
PI Kent RB, Ladner RC, Ley AC, Roberts BL, Markland W;
PI Guterman SK;
DR WPI; 92-331725/40.
DR N-PDSB; Q28836.
PT Developing new epitope(s) or binding proteins - using a display
PT phage library which is less biased and in which irreversible
PT binding to target is alleviated
PS Example 2; Page 124; 168pp; English.
CC The protein sequence is that translated from a construct for a
CC tripartite gene comprising the M13 major coat protein III signal
CC sequence, the bovine pancreatic trypsin inhibitor gene, and the
CC mature M13 major coat protein III gene. The construct was prep.
CC from 16 synthetic oligomers. Expression of the construct allows
CC expression of BPTI on the surface of M13 and alleviates problems
CC of irreversible binding of the display phage for the target and
CC also provides a less biased and hence more efficient phage library
CC to isolate high affinity epitopes or binding proteins.
CC See also R27083-96.
SQ Sequence 484 AA;

Query Match 2.6%; Score 7; DB 1; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 49 GGGSGSG 55
Db 328 GGGSGSG 334

RESULT 13
R35081
ID R35081 standard; Protein; 3080 AA.
AC R35081;
DT 28-MAY-1993 (first entry)
DE ZmV polyprotein.
KW Zucchini yellow mosaic virus; ZYMV; potyvirus; polyprotein; protease;
KW proteolytic activity; 49 kD protease; trypsin-like cysteine protease;
KW animal picornavirus; sissile bond; N1b; protein; coat.
OS Zucchini yellow mosaic virus.
FH Key Location/Qualifiers
FT cleavage_site 786..787
FT protein 1164..1165
/note= "Cleavage site between the aphid transmission
FT cleavage_site 1164..1165
/note= "Cleavage site between 46 kD protein and the

```

FT cytoplasmic inclusion protein (CI)"  
FT 1798. .1799  
FT /note= "Cleavage site between CI and VPg/protease (VPg  
FT and protease are probably not separated in  
FT zymv)." 2284. .2285  
FT /note= "Cleavage site between VPg/protease and RNA  
FT replicase (REP)" 2801. .2802  
FT /note= "Cleavage site between REP and the coat  
FT protein (Cp)"  
PN W09301305-A.  
PD 21-JAN-1993.  
PF 09-JUL-1992; U05745.  
PR 09-JUL-1991; US-727837.  
PA (BALI/) BALINT R.  
PI Balint R;  
WPI: 93-045506/05.  
N-PSDB; Q35297.  
PI Method for identifying protease inhibitors - useful for drugs  
PT screening for treating e.g. chronic inflammation, metastatic  
PT cancers and viral infections  
PS Disclosure: 62pp; English.  
CC "This sequence was not disclosed in the specification but was decoded  
CC (zymv) polyprotein. Zymv is a potyvirus and expresses its genome as a  
CC single 350 kD polyprotein which is cleaved into at least seven mature  
CC gene products by three distinct proteolytic activities. Two of the  
CC proteases are virus encoded, including the potyviral 49 kD protease.  
CC This protease is responsible for at least five of the seven cleavages.  
CC This enzyme is a trypsin-like cysteine protease which is structurally  
CC and mechanistically representative of the largest class of viral  
CC proteases, including those of the animal picornaviruses. This enzyme  
CC is highly specific and appears to recognise a region comprised of  
CC about seven amino acids surrounding the scissile bond. Of the five  
CC sites cleaved by this enzyme, the two flanking the protease appear to  
CC be cleaved intramolecularly, while the remaining three appear to be  
CC cleaved intermolecularly. Of the latter three, the site between the  
CC N1B protein and the coat protein appears to be the most active.  
SQ Sequence 3080 AA;

Query Match 2.6%; Score 7; DB 1; Length 3080;  
Best Local Similarity 100.0%; Pred. No. 6.9e-02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

51 GSGSGEK 57  
|||||||  
Db 2830 GSGSGEK 2836

RESULT 14  
R50174  
ID R50174 standard; Protein; 211 AA.  
AC R50174;  
DE 04-OCT-1994 (first entry)  
DE Membrane anchor peptide derived from M13 coat protein (gene III).  
KW Expression; insertion sequence; Vh chain; variable heavy chain;  
KW secretion signal; epitope tag; expression vector; lamda Zap;  
KW lamda Hc2; membrane; anchor peptide; bacteriophage; phage; M13;  
KW cpIII.  
OS Bacteriophage M13.  
PN W09405781-A.  
PD 17-MAR-1994.  
PF 03-SEP-1993; U08364.  
PR 04-SEP-1992; US-941369.  
PA (SCRI ) SCRIPPS RES INST.  
PI Lerner RA, Light JP;  
DR WPI: 94-101186/12.  
DR N-PSDB; Q44655.  
PT Filamentous phage comprising a heterologous polypeptide and a  
PT hetero:dimer - is used to detect the presence of a preselected  
PT ligand in a sample

PS Disclosure; Page 176-177; 232pp; English.  
CC A filamentous phage constructed so as to be able to detect the  
CC presence of a preselected ligand in a sample comprises (1) a  
CC heterologous polypeptide fused to a first filamentous phage coat  
CC protein membrane anchor and (2) a heterodimeric receptor  
CC comprising first and second receptor polypeptides, one of which  
CC is fused to a second filamentous phage coat protein membrane  
CC anchor. This sequence is a membrane anchor polypeptide derived  
CC from the M13 coat protein gene 3 (cpIII).  
SQ Sequence 211 AA;

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGSG 55  
|||||||  
Db 55 GGGSGSG 61

RESULT 15  
R54281  
ID R54281 standard; Protein; 211 AA.  
AC R54281;  
DE 10-NOV-1994 (first entry)  
DE M13 phage coat protein cpIII membrane anchor domain.  
KW Human immunodeficiency virus; HIV1; glycoprotein gp41; epitope;  
KW neutralisation; monoclonal antibody; light chain; variable region;  
KW filamentous phage M13; M13mpl8; coat protein cpIII.  
OS Bacteriophage M13.  
FT Key Location/Qualifiers  
FT Misc\_difference 183 /note= "no corresp. codon in Q64222"  
PN W09407922-A.  
PD 14-APR-1994.  
PF 30-SEP-1993; U09328.  
PR 30-SEP-1992; US-954148.  
PA (SCRI ) SCRIPPS RES INST.  
PI Barbas CF, Burton DR, Lerner RA;  
DR WPI: 94-135516/16.  
DR N-PSDB; Q64222.  
PT New human monoclonal antibodies neutralising HIV - react with  
PT gp120 or gp41 and nucleic acid encoding them, useful for in vivo  
PT or in vitro diagnosis and for passive immuno-therapy  
PS Example 1; Page 145-146; 248pp; English.  
CC M13mpl8 replicative form DNA was used as a template for isolating  
CC the gene encoding the membrane anchor domain at cpIII. Two PCR  
CC amplifications were performed for construction of a DNA fragment  
CC consisting of the mature gene for cpIII membrane anchor domain  
CC located 5' to a sequence encoding the lacZ promoter, operator and  
CC cap-binding site for controlling light chain expression.  
SQ Sequence 211 AA;

Query Match 2.6%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GGGSGSG 55  
|||||||  
Db 55 GGGSGSG 61

Search completed: February 11, 2000, 05:14:32  
Job time: 730 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 10, 2000, 13:27:02 ; Search time 506.51 Seconds  
(without alignments)  
6037.961 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTTAAAGGTC.....AAACCAATTTATATCCATAA 810

Scoring table: OLIGO\_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:  
B size : 0

Number of hits that pass the threshold : 9077268

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*
- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*

- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_gss1:\*
- 80: gb\_gss2:\*
- 81: gb\_gss3:\*
- 82: gb\_gss4:\*
- 83: em\_gss1:\*
- 84: em\_gss2:\*
- 85: em\_gss3:\*
- 86: em\_gss4:\*
- 87: gb\_gss5:\*
- 88: gb\_gss6:\*
- 89: gb\_gss7:\*
- 90: gb\_gss8:\*
- 91: gb\_gss9:\*
- 92: em\_gss5:\*
- 93: em\_gss6:\*
- 94: em\_gss7:\*
- 95: em\_gss8:\*
- 96: em\_gss9:\*
- 97: em\_gss10:\*
- 98: em\_gss11:\*
- 99: gb\_gss10:\*
- 100: gb\_gss11:\*
- 101: em\_gss12:\*
- 102: gb\_gss12:\*
- 103: gb\_gss13:\*
- 104: gb\_gss14:\*
- 105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	550	67.9	574	50	AI680904
2	505	62.3	556	47	AI479764
3	493	60.9	539	62	AI923178
4	479	59.1	537	48	AI580361
5	474	58.5	548	38	AA772225
6	456	56.3	507	50	AI682204
7	447	55.2	498	45	AI349520

- AI680904 tx42f05.x
- AI479764 tm69b04.x
- AI923178 wn67b10.x
- AI580361 tm47h02.x
- AA772225 ai41c01.s
- AI682204 wa71b06.x
- AI349520 qp72g05.x

8	441	54.4	530	50	A16744462	A16744462	wc44e01..x
9	436	54.2	439	44	A12991554	A12991554	qm96f11..x
10	439	53.8	471	43	A1168073	A1168073	oz89c02..x
11	423	52.2	423	44	A1299329	A1299329	qn22c06..x
12	395	48.8	395	44	A1245878	A1245878	qk27f08..x
13	389	48.0	448	41	A1038331	A1038331	ox84a06..x
14	368	45.4	500	50	A1680969	A1680969	tx35h07..x
15	320	39.5	322	62	A1903935	A1903935	MR-BT038-
16	320	39.5	398	64	AW071521	AW071521	xa36a07..x
17	319	39.4	323	62	A1903938	A1903938	MR-BT038-
18	319	39.4	323	62	A1903940	A1903940	MR-BT038-
19	311	38.4	507	25	N47594	N47594	yy59a07..s1
20	306	37.8	632	41	A1057115	A1057115	oz23g10..x
21	301	37.2	335	43	A1194064	A1194064	qe78d08..x
22	281	34.7	461	50	A1693751	A1693751	wd44h04..x
23	280	34.6	392	38	AA757694	AA757694	wz38h03..s
24	277	34.2	382	32	AA345598	AA345598	EST51648
25	267	33.0	385	25	D81362	D81362	HUM159G08B
26	257	31.7	300	62	A1887092	A1887092	w196f10..x
27	250	30.9	508	36	AA639448	AA639448	nq88g05..s
28	242	29.9	252	24	N36461	N36461	yx83e05..r1
29	226	27.9	540	44	A1284315	A1284315	qj65c02..x
30	214	26.4	464	24	N34957	N34957	yy50b11..s1
31	202	24.9	590	69	AW140060	AW140060	UI-H-BT1-
32	201	24.8	580	23	R99199	R99199	yq64d10..r1
33	185	22.8	674	29	AA134062	AA134062	zl49h02..s
34	182	22.5	416	38	AA757905	AA757905	zg42h09..s
35	172	21.2	474	43	A1239974	A1239974	qh44d02..x
36	169	20.9	305	40	AA927408	AA927408	om27d10..s
37	165	20.4	216	30	AA253249	AA253249	tr53a03..r
38	165	20.4	565	46	A1420901	A1420901	tf03g10..x
39	160	19.8	475	31	AA306979	AA306979	EST178050
40	160	19.8	482	51	A1732252	A1732252	ye35c09..x
41	148	18.3	346	22	R26459	R26459	yh50e12..s1
42	140	17.3	480	24	N28591	N28591	yx38c03..r1
43	129	15.9	351	31	AA302858	AA302858	EST12947
44	122	15.1	417	23	D61247	D61247	HUM172D06B
45	113	14.0	450	41	A1066635	A1066635	oz82c10..x

## ALIGNMENTS

RESULT 1  
 AI680904  
 LOCUS  
 DEFINITION  
 tx142f05.x1 NCI-CGAP.Lu24 Homo sapiens cDNA clone IMAGE:2272257 3'  
 nt42f05 to WP:02F5.3 CE00039 GTP-BINDING PROTEIN;; mRNA sequence.  
 ACCESSION--AI680904  
 VERSION--  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 574)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 COMMENT  
 On Dec 20, 1995, this sequence version replaced gi:134035.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.11nl.gov/bbrp/image/image.html

<b>FEATURES</b>					
Seq primer: -40UP from Gibco					
High quality sequence stop: 459.					
Location/Qualifiers					
1. 574					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="IMAGE:2272257"					
/clone_lib="NCI_CGAP_Lu24"					
/tissue_type="carcinoid"					
/lab_host="DH10B"					
/note="organ: lung; Vector: pT7 <sup>+</sup> 3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1523439). Subtraction by Bento Soares and M. Fatima Bonaldo."					
BASE COUNT      148 a    118 c    136 g    172 t					
ORIGIN					
Query Match                  67.9%; Score 550; DB 50; Length 574;					
Best Local Similarity 100.0%; Pred. No. 1.7e-288;					
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	235	GCAGACTCGTTGGTGCTCCTGTGGTCGCTCAGTCACTACAGGACCCTGGGGGCGTGTT	294		
DB	25	GCAGACTCGTTGGTGCTCCTGTGGTCGCTCAGTCACTACAGGACCCTGGGGGCGTGTT	84		
QY	295	GCCACCTCCGCCGGGGCGAGGAGTGCCTTAAGTGCAGGACCTCAAAGTGGACAATAT	354		
DB	85	GCCACCTCCGCCGGGGCGAGGAGTGCCTTAAGTGCAGGACCTCAAAGTGGACAATAT	144		
QY	355	ATTGTGAAGATCCAAAAATAAATGACGCTACGCAAGAACCATTTAACTGTACAACTAC	414		
DB	145	ATTGTGAAGATCCAAAAATAAATGACGCTACGCAAGAACCATTTAACTGTACAACTAC	204		
QY	415	ACAGCTCATGTTTCTGTTTTCCAGCACCAACATACTTGTAAGGATTCCTCAGTGGCAAT	474		
DB	205	ACAGCTCATGTTTCTGTTTTCCAGCACCAACATACTTGTAAGGATTCCTCAGTGGCAAT	264		
QY	475	GAACACATTTTACTGGGACGAAGTTGGTTTTTTCAGCCCCATATCTTGCCGAAATGTA	534		
DB	265	GAACACATTTTACTGGGACGAAGTTGGTTTTTTCAGCCCCATATCTTGCCGAAATGTA	324		
QY	535	AATGGCTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTTGATGGTTGGGAGCA	594		
DB	325	AATGGCTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTTGATGGTTGGGAGCA	384		
QY	595	GATCGATTTTACTTGGATACCTGCTTTGGGTTGCTTAAAGTTTTCACGTAGGTTT	654		
DB	385	GATCGATTTTACTTGGATACCTGCTTTGGGTTGCTTAAAGTTTTCACGTAGGTTT	444		
QY	655	TGTGGAATGGGAGCCTAATTGATTTCACTCTATTTCATCGAGATGTTGGACCTTCA	714		
DB	445	TGTGGAATGGGAGCCTAATTGATTTCACTCTATTTCATCGAGATGTTGGACCTTCA	504		
QY	715	GATGGAGTAGTTACATTTATAGATTACTATGAACACAGACTTACAAAGCTAGTATTACT	774		
DB	505	GATGGAGTAGTTACATTTATAGATTACTATGAACACAGACTTACAAAGCTAGTATTACT	564		
QY	775	AATGAACAT 784			
DB	565	AATGAACAT 574			
RESULT     2					
AI479764					
LOCUS					
tm69b04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163343 ,					
DEFINITION					
mRNA sequence.					
EST                  14-APR-1999					

```

ACCESSION   AI479764
VERSION     AI479764.1  GI:4372932
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 356)
AUTHORS    NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/RTGAP), Tumor Gene Index
            Unpublished (1998)
JOURNAL     On Jun 5, 1998 this sequence version replaced gi:3188819.
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 614 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2163343"
/clone_lib="NCI-CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCCGATAGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT  148 a 109 c 130 g 168 t 1 others
GIN

Query Match      62.3%; Score 505; DB 47; Length 556;
Best Local Similarity 99.8%; Pred. No. 5.4e-264;
Matches 555; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 234 GGCACACTCGTGTGGTCTGCTGGTTCGTCAGTCACACAGACCTGGGGGCTGT 293
DB 1 GGCACACTCGTGTGGTCTGCTGGTTCGTCAGTCACACAGACCTGGGGGCTGT 60
QY 294 TGCACCTCCGGGGGGGAGAGTCGGTTAAGTCGAGGACCTCAAAAGTGGACAATA 353
DB 61 TGCACCTCCGGGGGGGAGAGTCGGTTAAGTCGAGGACCTCAAAAGTGGACAATA 120
QY 354 TATTGTGAAGATCCAAAATAAATGACGCTACGCAAGAACACAGTTAACTGTACAAACTA 413
DB 121 TATTGTGAAGATCCAAAATAAATGACGCTACGCAAGAACACAGTTAACTGTACAAACTA 180
QY 414 CACAGCTCATGTTCTCTGTTTCCAGCACCCACCAATACTGTGAAGATCCAGTGGCAA 473
DB 181 CACAGCTCATGTTCTCTGTTTCCAGCACCCACCAATACTGTGAAGATCCAGTGGCAA 240
QY 474 TGAACACACATTTTACTGGGAACGAAGTGTGTTTTTTCAGCCCATATCTTGGCGAAATGT 533

```

us-09-060-609-1\_1.lst

Db 241 TGAACACACATTTTACTGGGAACGAAGTGTGTTTTTTCAGCCCATATCTTCCGCAATGT 300

QY 534 AAATGGCTATTCTCTACAAAGTGGCAGTCGCATTCCTCTTTTCTTGGATGTTGGGAGC 593

Db 301 AAATGGCTATTCTCTACAAAGTGGCAGTCGCATTCCTCTTTTCTTGGATGTTGGGAGC 360

QY 594 AGATCGATTTTACTCTGGATACCCCTGCTTTGGGTTTGTAAAGTTTTCACATGTAGGTT 653

Db 361 AGATCGATTTTACTCTGGATACCCCTGCTTTGGGTTTGTAAAGTTTTCACATGTAGGTT 420

QY 654 TTGTGAATTTGGGAGCCCTAATTGATTTTCATCTTATTCATGCAGATGTTGGACCTTC 713

Db 421 TTGTGAATTTGGGAGCCCTAATTGATTTTCATCTTATTCATGCAGATGTTGGACCTTC 480

QY 714 AGATGAAGTAGTTACATTATAGATTACTATGGAACACAGACTTCAAGACATGAGTATTAC 773

Db 481 AGATGAAGTAGTTACATTATAGATTACTATGGAACACAGACTTCAAGACATGAGTATTAC 540

QY 774 TAATGAACATTTAGA 789

Db 541 TAATGAACATTTAGA 556

RESULT 3

AI923178

LOCUS

DEFINITION

AI923178 599 bp mRNA EST 02-SEP-1999

similar to WP:02F5.3 CE00039 GTP-BINDING PROTEIN ; mRNA sequence.

ACCESSION

AI923178

VERSION

AI923178.1 GI:5659142

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 599)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

On Dec 20, 1995 this sequence version replaced gi:1133491.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 457.

Location/Qualifiers

1..599

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2450491"

/clone\_lib="NCI-CGAP\_Lu19"

/tissue\_type="squamous cell carcinoma, poorly

differentiated (4 pooled tumors, including primary and

metastatic)"

/dev\_stage="adult"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from

pooled lung tumor tissue, and was then primed with a Not I

- oligo(dT) primer. Double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization. Library constructed by Bento Soares and M. Fatima Ronaldo.

BASE COUNT		ORIGIN	
148 a	127 c	146 g	176 t
		2 others	

Query Match	60.98;	Score 493;	DB 62;	Length 599;
Best Local Similarity	99.78;	Prod NO	1	88-257.

Qy 190 GCGGCCGCGCTGCCCGTCTGTCCGTCCTCCGAGGCCGTGACGCCACACTCGTTGGT 249

Db 1 GCGGCCGCTGGCCGTCTGGTCCGTCTGCTCCGAGGCCGTGACGCCAGACTCGTTGGT 60

QY 250 GTCTGTGGTTCTGTTCTCAGTCACTACAGGACCCCTGGGGGGCTGTTGCCACCTCCGCCGGG 309

Db 61 GTCTGTGGTTGCTCAGTCACTACAGGACCTGGGGGGCTGTGCCACCTCCGCGGG 120

310 GCGGAGGAGTCGCTTAAGTGGAGGACCTCAAAGTGGGCAATATATTGTGAAGATCCA 369

Db 121 GCGAGGAGTCGCTTAAGTGCAGGACCTCAAGTGGACAATATATTGTGAAGATCCA 180

QY 370 AAAATAATGACGGTACGCCAGAACCAAGTTAACTGTACAAACTACACAGCTCATGTTCC 429

Db 181 AAAATAATGACGCTACGCCAGAACCACTTAAGTGTACAAACTACACAGCTCATGTTCC 240

430 TGTTTTCCAGCACCACATACACTTGTGAAGGATCCAGTGGCAATCAAACACATTTTACT 489

D5 241 TGTTTTCCAGCACCACATCACTTGTAAGGATTCCAGTGGCAATGAACACATTCT 300

QY 490 GGSAAACGAAGTTGGTTTTTTTCAAGCCCATATCTTCCGGAATGFAAATGGCTATTCCTAC 549

D6 301 GGGAAACGAAGTGGGTTTTTTTCAAGCCCATATCTTGCCGAAATGTAAAGGGCTATTCTTAC 360

609  
 610  
 611  
 612  
 613  
 614  
 615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633  
 634  
 635  
 636  
 637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647  
 648  
 649  
 650  
 651  
 652  
 653  
 654  
 655  
 656  
 657  
 658  
 659  
 660  
 661  
 662  
 663  
 664  
 665  
 666  
 667  
 668  
 669  
 670  
 671  
 672  
 673  
 674  
 675  
 676  
 677  
 678  
 679  
 680  
 681  
 682  
 683  
 684  
 685  
 686  
 687  
 688  
 689  
 690  
 691  
 692  
 693  
 694  
 695  
 696  
 697  
 698  
 699  
 700  
 701  
 702  
 703  
 704  
 705  
 706  
 707  
 708  
 709  
 710  
 711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755  
 756  
 757  
 758  
 759  
 760  
 761  
 762  
 763  
 764  
 765  
 766  
 767  
 768  
 769  
 770  
 771  
 772  
 773  
 774  
 775  
 776  
 777  
 778  
 779  
 780  
 781  
 782  
 783  
 784  
 785  
 786  
 787  
 788  
 789  
 790  
 791  
 792  
 793  
 794  
 795  
 796  
 797  
 798  
 799  
 800  
 801  
 802  
 803  
 804  
 805  
 806  
 807  
 808  
 809  
 810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841  
 842  
 843  
 844  
 845  
 846  
 847  
 848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863  
 864  
 865  
 866  
 867  
 868  
 869  
 870  
 871  
 872  
 873  
 874  
 875  
 876  
 877  
 878  
 879  
 880  
 881  
 882  
 883  
 884  
 885  
 886  
 887  
 888  
 889  
 890  
 891  
 892  
 893  
 894  
 895  
 896  
 897  
 898  
 899  
 900  
 901  
 902  
 903  
 904  
 905  
 906  
 907  
 908  
 909  
 910  
 911  
 912  
 913  
 914  
 915  
 916  
 917  
 918  
 919  
 920  
 921  
 922  
 923  
 924  
 925  
 926  
 927  
 928  
 929  
 930  
 931  
 932  
 933  
 934  
 935  
 936  
 937  
 938  
 939  
 940  
 941  
 942  
 943  
 944  
 945  
 946  
 947  
 948  
 949  
 950  
 951  
 952  
 953  
 954  
 955  
 956  
 957  
 958  
 959  
 960  
 961  
 962  
 963  
 964  
 965  
 966  
 967  
 968  
 969  
 970  
 971  
 972  
 973  
 974  
 975  
 976  
 977  
 978  
 979  
 980  
 981  
 982  
 983  
 984  
 985  
 986  
 987  
 988  
 989  
 990  
 991  
 992  
 993  
 994  
 995  
 996  
 997  
 998  
 999  
 1000

[illegible]

q7	GGAATACCCCTGTTTGGGTTTGTAATAAGTTTTGCATGTATGGGGTTTTTGCGAATTGGGAGC
o19	GGATATTCCTGTTTGGGTTTGTAATAAGTTTTGCATGTATGGGGTTTTTGCGAATTGGGAGC
nb	CGATACCCTGTTTGGGTTTGTAATAAGTTTTGCATGTATGGGGTTTTTGCGAATTGGGAGC

[illegible]

481 CTAATTGATTTCATCTCTATTCAATCAGATTGTTGGACCTTCACATGGAGTACGTAC 540  
 |||||  
 481 CTAATTGATTTCATCTCTATTCAATCAGATTGTTGGACCTTCACATGGAGTACGTAC 540  
 |||||

730 ATTATAGANTTACTATGGAAACGACCTTACAGACTGAGTANTTACTAATGAAACAT 784

Db 541 ATTATAGATTACTATGGAACACAGACTTCAAGAGACTGAGTATTACTAATGAACAT 595

LOCUS	AI580361	537 bp	mrna	EST	05-APR-1999
DEFINITION	tm47h02.x1	NCT CGAP Kid1	Homo sapiens	cdna clone	IMAGE:216120837

similar to WP: C02F5.3 CE00039 GTP-BINDING PROTEIN ; mRNA sequence  
AI580361

VERSION AI580361.1 GI:4564737  
KEYWORDS EST.

SOURCE	ORGANISM
human.	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 537)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

**TITLE**  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949103.

Contact: Robert Strauss, Ph.D.  
Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

AA772225 548 bp mRNA EST 31-DEC-1998  
 LOCUS a14ic01.s1 Soares\_parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone  
 DEFINITION 1359552 3' similar to WP:02F5.3 CE00039 GTP-BINDING PROTEIN ;  
 mRNA sequence.  
 AA772225  
 ACCESSION AA772225.1 GI:2824008  
 VERSION  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 548)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1995 this sequence version replaced gi:811076.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: similarity on wrong strand  
 Possible reversed clone: polyt not found  
 Insert Length: 978 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 459.

## FEATURES

Location/Qualifiers  
 1..548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="1359552"  
 /clone\_lib="Soares\_parathyroid\_tumor\_NbHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: parathyroid gland; Vector: pT7R3D  
 (Pharmacia) with a modified polylinker; Site\_1: Not I;  
 Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer  
 (5'-  
 TGTTACCAATCTGAAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTT  
 T-3')], double-stranded cDNA was size selected, ligated to  
 Eco RI adapters (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of a modified pT7R3  
 vector (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
 adenomas was kindly provided by Dr. Stephen Marx, National  
 Institute of Diabetes and Digestive and Kidney Diseases,  
 NIH."

BASE COUNT 146 a 106 c 126 g 169 t 1 others  
 ORIGIN

Query Match 58.5%; Score 474; DB 38; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 4e-247;  
 Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 315 GGAGTCGCTTAAGTCGAGGACCTCAAGTGGGACATATATTGTAAACATCCAAAT 374  
 Db 75 GGAGTCGCTTAAGTCGAGGACCTCAAGTGGGACATATATTGTAAAGATCCAAAT 134  
 Qy 375 AAATGACGCTACGCAAGAACCACTGTAACACTGACAACTACACAGCTCAATGTTT 434  
 Db 135 AAATGACGCTACGCAAGAACCACTGTAACACTGACAACTACACAGCTCAATGTTT 194

Qy 435 TCAGACACCCCAACATAACTTGTAAAGATTCCAGTGGCAATGAACACATTTTACTGGAA 494  
 Db 195 TCAGACACCCCAACATAACTTGTAAAGATTCCAGTGGCAATGAACACATTTTACTGGAA 254  
 Qy 495 CGAAGTGGTTTTTCAAGCCCATATCTTCCCAAAATGTAATGGCTATTCCTACAAAGT 554  
 Db 255 CGAAGTGGTTTTTCAAGCCCATATCTTCCCAAAATGTAATGGCTATTCCTACAAAGT 314  
 Qy 555 GGCAGTCGCATCTCTCTTTTCTTGGATGTTGGAGCAGATCGATTATTTACCTTGGATA 614  
 Db 315 GGCAGTCGCATCTCTCTTTTCTTGGATGTTGGAGCAGATCGATTATTTACCTTGGATA 374  
 Qy 615 CCTGCTTTGGGTTTGTAAAGTTTTCAGTGTGGACCTTCAGATGGAAGTGGAGCCCTAAT 674  
 Db 375 CCTGCTTTGGGTTTGTAAAGTTTTCAGTGTGGACCTTCAGATGGAAGTGGAGCCCTAAT 434  
 Qy 675 TGATTTCACTTCTTATTTCATATGCAATGTTGGACCTTCAGATGGAAGTGGAGCTTACATAT 734  
 Db 435 TGATTTCACTTCTTATTTCATATGCAATGTTGGACCTTCAGATGGAAGTGGAGCTTACATAT 494  
 Qy 735 AGATTACTATGGAACACAGACTTACAAGCTGAGTATTACTTAATCAACATTTAG 788  
 Db 495 AGATTACTATGGAACACAGACTTACAAGCTGAGTATTACTTAATCAACATTTAG 548

## RESULT 6

LOCUS AI682204 507 bp mRNA EST 26-MAY-1999  
 DEFINITION wa7lb06.xl Soares\_NFL\_T\_GBC.S1 Homo sapiens cDNA clone  
 IMAGE:2301587 3', mRNA sequence.  
 ACCESSION AI682204  
 VERSION AI682204.1 GI:4892386  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 507)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1136077.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 469.

## FEATURES

Location/Qualifiers  
 1..507  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2301587"  
 /clone\_lib="Soares\_NFL\_T\_GBC.S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT7R3D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBH19W, testis NHT, and B-cell  
 NCI-CGAP-GCBI) were mixed, and ss circles were used as  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 120 a 109 c 125 g 152 t 1 others

## BASE COUNT

ORIGIN



Query Match	56.3%	Score 456;	DB 50;	Length 507;
Best Local Similarity	99.9%	Pred. No. 2.5e-237;		
Matches 506;	Conservative	0; Mismatches 1;	Indels 0;	Gaps 0;
QY 208	GGTCCGCTCCTCGGAGGCGGTACAGCCAGACTCGTTGGTGTCTCTGTTGGTTCGTCTCA	267		
DB 1	GGTCCGCTCCTCGGAGGCGGTACAGCCAGACTCGTTGGTGTCTCTGTTGGTTCGTCTCA	60		
QY 268	GTCACTACAGACCCCTGGGGGCGTGTGCCACCTCCCGCGGGGCGAGAGTCGCTTAAG	327		
DB 61	GTCACTACAGACCCCTGGGGGCGTGTGCCACCTCCCGCGGGGCGAGAGTCGCTTAAG	120		
QY 328	TGGGAGGACCTCAAGTGGGACAATATATTGTAAAGATCCAAAATAAATGACGCTACG	387		
DB 121	TGGGAGGACCTCAAGTGGGACAATATATTGTAAAGATCCAAAATAAATGACGCTACG	180		
QY 388	CAGAACCACTTAACTACAAACTACACACTCATGTTTCCTGTTTCCAGCACCCCAAC	447		
DB 181	CAGAACCACTTAACTACAAACTACACACTCATGTTTCCTGTTTCCAGCACCCCAAC	240		
QY 448	ATACTTGTAAAGATTCAGTGGCAATGAACACATTTTACTGGGAACGAAGTTGGTTTT	507		
DB 241	ATACTTGTAAAGATTCAGTGGCAATGAACACATTTTACTGGGAACGAAGTTGGTTTT	300		
QY 508	TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCTCTACAAAGTGGCAGTCGCAFTG	567		
DB 301	TTCAAGCCCATATCTTGCCGAAATGTAAATGGCTATTCTCTACAAAGTGGCAGTCGCAFTG	360		
QY 568	TCCTCTTTTCTTGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTTGGGT	627		
DB 361	TCCTCTTTTCTTGATGGTTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTTGGGT	420		
QY 628	TTGTAAAGTTTTGCACCTGTAGGGTTTTGTGGAATTTGGGAGCCCAATTTGATTTCAATCTT	687		
DB 421	TTGTAAAGTTTTGCACCTGTAGGGTTTTGTGGAATTTGGGAGCCCAATTTGATTTCAATCTT	480		
QY 688	ATTTCATGCAGATGTTGGACCTTCA	714		
DB 481	ATTTCATGCAGATGTTGGACCTTCA	507		

```

/lab_host="DH10B (ampicillin resistant)"
/notes=Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fátima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbH819W."
113 a 112 c 128 g 144 t 1 others
. BASE COUNT

```

BASE COUNT	113 a	112 c	128 g	144 t	1 others
ORIGIN					

Query Match	55.28;	Score 447;	DB 45;	Length 498;
Best Local Similarity	99.88;	Pred. No. 2e-232;		

[illegible]

QY 250 GTCCTGTGGTTCGTCACGTCACTACAGGACCCCTGGGGGGCTGTGGCACCTCCCGCGGG 309

Db 61 GTCCTGTGGTCTCAGTCACCTACGAGACCTGGGGGGCTGTGGCCACCTCCCGCGGG 120

Qy 310 GCGGAGGAGTCGCTTAAAGTGGAGGACCTCAAAGTGGGCAATATATTTGTAAGATCCA 369

Db 121 GCGAGAGTCGCTTAAGTGGAGGACCTCAAGTGGACAATATATTGTAAAGATCCA 180

Qy 370 AAAATAAATGACGCTACGCAAGAACAGTTAACTGTGCAAAACTACACAGCTCATGTTTC 429

Db 181 AAAATAATGACCTACGCAGAACCGATTAACTGTACAAACTACACAGCTCATGTTCC 240

	TGTTTCCAGCACCACCAAT	AACTCTTAAGGAATTCCAGTGCGAA	TGAAAACACATTTTTACT	489
Qy				
Dk	MCMNMNMC	CACGCACA	GMMNCMA	360

DD 161111CCAGGACCCCAACATATC1161AAGGAT1CCAGTGGCAATGAATACACATTTTAC1 300  
 241  
 QV 490 GGGACGCAAGTTGGTTTTTCAAGCCCATATCTTGGCGAAATGTAATGGCTATTCTTAC 549

Db 301 GGGACGAAGTTGGTTTTTCAAGCCCATATCTTGGCGAATGTAATGGCTATTCTTAC 360

QY 550 AAAGTGCAGTCGCATGTCTCTTTTCTTGGATGGTTGGGAGCAGATCGATTTTACCTT 609

361 AAAGTGGCAGTCGCATTGTCCTTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTT 420

QY 610 GGATACCCCTGCTTTGGGTTTGTTAAAGTTTTGCAC<sup>1</sup>TGTAGGGTTTTCTCGAATGGGAGC 669

Db 421 GGATACCTGCTNTGGTTTCTTAAAGTTTTCACATGTAGGTTTTCGGAAATGGGAGC 480

Qy 670 CTAATTCATTTCATTCTT 687

D6 481 CTAATTGATTTCATTCTT 498

RESULT 8  
AI674462

LOCUS	AI674462	530 bp	mRNA	EST	19-MAY-1999
DEFINITION	wc44e01.x1 NCI_CGAP_Pr28	Homo sapiens	CDNA clone	IMAGE:2321496	3'

similar to WP\_C02F5.3 CE00039 GTP-BINDING PROTEIN ; mRNA sequence.  
ACCESSION AI674462

VERSION AI674462.1 GI:4874942  
KEYWORDS EST.  
CONTACT 1-800-222-2222

ORGANISM	Source
Homo sapiens	Human
Fukuryuina	Metazoa
Craniata	Chordata
Vertebrata	Mammalia

REFERENCE 1 (bases 1 to 530)

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL  
COMMENT  
Tumor Gene Index  
Unpublished (1997)  
On Mar 10, 1998 this sequence version replaced gi:2948555.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 459.  
FEATURES  
source  
1..530  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2321496"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 123 a 116 c 133 g 157 t 1 others  
ORIGIN  
|||||

Query Match 54.4%; Score 441; DB 50; Length 530;  
Best Local Similarity 99.8%; Pred. No. 3.7e-229;  
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 232 ACGGCAGACTCGTTGGTCTCTGTTGCTCTAGTCAGTCAGTCAGGACCCCTGGGGGGCT 291  
Db 39 ACGGCAGACTCGTTGGTCTCTGTTGCTCTAGTCAGTCAGTCAGGACCCCTGGGGGGCT 98  
QY 292 GTTGCACCTCCGCGGGGGGAGGAGTCGCTTAAGTCGAGGACCTCAAAGTGGGACAA 351  
99 GTTGCACCTCCGCGGGGGGAGGAGTCGCTTAAGTCGAGGACCTCAAAGTGGGACAA 158  
QY 352 TATATTGTAAGATCCAAAATAATGACGCTACGTCAGCAGACACCAGTTAACTGTACAAC 411  
Db 159 TATATTGTAAGATCCAAAATAATGACGCTACGTCAGCAGACACCAGTTAACTGTACAAC 218  
QY 412 TACACAGCTCATGTTCTCTGTTTCCAGCACCAACATAACTGTGAAGATTCCAGTGGC 471  
Db 219 TACACAGCTCATGTTCTCTGTTTCCAGCACCAACATAACTGTGAAGATTCCAGTGGC 278  
QY 472 AATGAACACATTTTACTGGGAACAGTGGTTTTTCAAGCCCATATCTTCCCGAAAT 531  
Db 279 AATGAACACATTTTACTGGGAACAGTGGTTTTTCAAGCCCATATCTTCCCGAAAT 338  
QY 532 GTAATGGCTATCTTACAAAGTGGCAGTCGATGTCCTCTTTTCTTGATGGTTGGGA 591  
Db 339 GTAATGGCTATCTTACAAAGTGGCAGTCGATGTCCTCTTTTCTTGATGGTTGGGA 398  
QY 592 GCAGATCGATTTTACCTGGATACCTGCTTGGGTTTCTTAAAGTTTTCACATGTAGG 651  
Db 399 GCAGATCGATTTTACCTGGATACCTGCTTGGGTTTCTTAAAGTTTTCACATGTAGG 458  
QY 652 TTTTGTGAATGGGACCTTAATTGATTCATCTTATTTCAATGAGATTGTTGGACCT 711  
|||||

Db 459 TTTGTGGAATTGGAGCCCTAATTGATTTCATTCTTATTTCAATGCAGATTGTTGGACCT 518  
QY 712 TCAGATGGAAGT 723  
|||||  
Db 519 TCAGATGGAAGT 530  
RESULT 9  
AI299154  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AI299154 439 bp mRNA EST 29-JAN-1999  
gm96f11.xl NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1896621 3',  
mRNA sequence.  
AI299154  
AI299154.1 GI:3958808  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 439)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 966 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 419.  
FEATURES  
source  
1..439  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1896621"  
/clone\_lib="NCI\_CGAP\_Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 105 a 100 c 110 g 124 t  
ORIGIN  
|||||

Query Match 54.2%; Score 439; DB 44; Length 439;  
Best Local Similarity 100.0%; Pred. No. 4.6e-228;  
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 208 GGTCCGTCTCTCCGAGGCCGTGACGCCAGACTCGTTGGTGTCTGTGCTCTCTCA 267  
Db 1 GGTCCGTCTCTCCGAGGCCGTGACGCCAGACTCGTTGGTGTCTCTGTCTCTCA 60  
QY 268 GTCACTACAGACCTCTGGGGGGCTGTTGCCACCTCCCGCGGGGGGAGGAGTCGCTTAAG 327  
Db 61 GTCACTACAGACCTCTGGGGGGCTGTTGCCACCTCCCGCGGGGGGAGGAGTCGCTTAAG 120  
QY 328 TGGGAGGACCTCAAGTGGGACATATATTTGTAAGATCCAAAATAAATGACGCTAGC 387  
|||||

Db 121 TCGAGGACCTCAAGTGGGCAATATATTGTAAAGATCCAAAATAAATGACGCTACG 180  
QY 388 CAAGAACCACTTAAGTGTACAACTACACAGCTCATGTTTCTGTTTCCAGCACCCCAAC 447  
Db 181 CAAGAACCACTTAAGTGTACAACTACACAGCTCATGTTTCTGTTTCCAGCACCCCAAC 240  
QY 448 ATAACTGTAAAGTTCAGTGGCAATGAACACATTTTACTGGAGCAAGATTGGTTTT 507  
Db 241 ATAACTGTAAAGTTCAGTGGCAATGAACACATTTTACTGGAGCAAGATTGGTTTT 300  
QY 508 TTCAAGCCCATATCTTCGGAATGTAATGGCTATCTTCAAGTGGAGTGGCATG 567  
Db 301 TTCAAGCCCATATCTTCGGAATGTAATGGCTATCTTCAAGTGGAGTGGCATG 360  
QY 568 TCTCTTTTCTTGGATGTTGGGAGCAGATCGATTACCTTGGATACCCCTGCTTTGGGT 627  
Db 361 TCTCTTTTCTTGGATGTTGGGAGCAGATCGATTACCTTGGATACCCCTGCTTTGGGT 420  
628 TTGTTAAAGTTTGCACGTG 646  
Db 421 TTGTTAAAGTTTGCACGTG 439

## RESULT 10

AI168073 471 bp mRNA EST 05-OCT-1998  
LOCUS oz89c02.x1 Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA  
DEFINITION clone IMAGE:1682498 3', mRNA sequence.  
ACCESSION AI168073  
VERSION AI168073.1 GI:3701243  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 471)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2282128.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

source  
1..471  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1682498"  
/tissue\_type="senescent\_fibroblasts\_NBHSF"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker V.TYE: phagemid; Site.1: Not I; Site.2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5'  
TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo."  
BASE COUNT 111 a 110 c 120 g 130 t  
ORIGIN  
Query Match 53.8%; Score 436; DB 43; Length 471;  
Best Local Similarity 100.0%; Pred. No. 2e-226;  
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 GCGGCGCCTGGCCGCTGGTCCGTCCTCGGAGGCGGTGACGCCACAGACTCGTTGGT 249  
Db 1 GCGGCGCCTGGCCGCTGGTCCGTCCTCGGAGGCGGTGACGCCACAGACTCGTTGGT 60  
QY 250 GTCCTGTGGTTCCTCAGTCACTACAGGACCTGGGGGGCTGTTGCCACCTCCGCCGGG 309  
Db 61 GTCCTGTGGTTCCTCAGTCACTACAGGACCTGGGGGGCTGTTGCCACCTCCGCCGGG 120  
QY 310 GCGAGGAGTGCCTTAAGTGGAGGACCTCAAAGTGGGCAATATATTGTTAAAGATCCA 369  
Db 121 GCGAGGAGTGCCTTAAGTGGAGGACCTCAAAGTGGGCAATATATTGTTAAAGATCCA 180  
QY 370 AAAATAAATGACGCTACGCAAGAACCACTTAAGTGGGCAATATATTGTTAAAGATCCA 429  
Db 181 AAAATAAATGACGCTACGCAAGAACCACTTAAGTGGGCAATATATTGTTAAAGATCCA 240  
QY 430 TGTTCCTCAGCACCACATCACTTGTAAAGTGGGCAATATATTGTTAAAGATCCA 489  
Db 241 TGTTCCTCAGCACCACATCACTTGTAAAGTGGGCAATATATTGTTAAAGATCCA 300  
QY 490 GGAACGAAGTTGTTTTTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 549  
Db 301 GGAACGAAGTTGTTTTTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 360  
QY 550 AAAGTGGCAGTCGCATGCTCTTTTCTTGGATGTTGGGAGCAGATCGATTTTACCTT 609  
Db 361 AAAGTGGCAGTCGCATGCTCTTTTCTTGGATGTTGGGAGCAGATCGATTTTACCTT 420  
QY 610 GGATACCTGCTTTGG 625  
Db 421 GGATACCTGCTTTGG 436

## RESULT 11

AI299329 423 bp mRNA EST 01-FEB-1999  
LOCUS qn22c06.x1 NCI-CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1898986 3',  
DEFINITION mRNA sequence.  
ACCESSION AI299329  
VERSION AI299329.1 GI:3958983  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 423)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA sequencing by: Greg Lennon, Ph.D.  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1021 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 410.  
Location/Qualifiers  
1..423  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1898986"  
/clone\_lib="NCI-CGAP\_Lu5"  
/tissue\_type="carcinoid"

## FEATURES

source  
1..423  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1898986"  
/clone\_lib="NCI-CGAP\_Lu5"  
/tissue\_type="carcinoid"



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 448)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 9, 1995 this sequence version replaced gi:802984.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1099 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 432.  
Location/Qualifiers  
1..448  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares-senescent-fibroblasts\_NbHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT7T3D (Pharmacia) with a modified  
polylinker V-type: phagemid; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTTACCAATCTGAAGTGGGCGCGCATTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 118 a 92 c 105 g 133 t  
ORIGIN

Query Match 48.0%; Score 389; DB 41; Length 448;  
Best Local Similarity 100.0%; Pred. No. 7.5e-201;  
Matches 389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 260 TCGTCTCAGTACAGGACCTCGGGGGCTGTTGCCACCTCCGCGGGGCGAGGAGT 319  
3 TCGTCTCAGTACAGGACCTCGGGGGCTGTTGCCACCTCCGCGGGGCGAGGAGT 62  
320 CGCTTAAGTGGGAGACCTCAAAAGTGGGACATATATTGTGAAGATCCAAAATAATG 379  
63 CGCTTAAGTGGGAGACCTCAAAAGTGGGACATATATTGTGAAGATCCAAAATAATG 122  
QY 380 ACGCTACGCAAGACCACTTAACCTACAACTACAGCTCATCTTCCTGTTTCCAG 439  
DB 123 ACGCTACGCAAGACCACTTAACCTACAACTACAGCTCATCTTCCTGTTTCCAG 182  
QY 440 CACCCACATCACTTGTAAAGATTCAGTGGCAATGAACACATTTTACTGGGAACGAAG 499  
DB 183 CACCCACATCACTTGTAAAGATTCAGTGGCAATGAACACATTTTACTGGGAACGAAG 242  
QY 500 TTGTTTTTCAAGCCCATATCTTCCGGAATGTAAATGGCTATTCCATACAAAGTGGCAG 559  
DB 243 TTGTTTTTCAAGCCCATATCTTCCGGAATGTAAATGGCTATTCCATACAAAGTGGCAG 302  
QY 560 TCGCATTTGCTCTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTTGGATACCCCTG 619  
DB 303 TCGCATTTGCTCTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTTGGATACCCCTG 362  
QY 620 CTTTGGGTTTGTAAAGTTTTCACCTGTA 648  
DB 363 CTTTGGGTTTGTAAAGTTTTCACCTGTA 391

RESULT 14

AI680969 500 bp mRNA EST 26-MAY-1999  
LOCUS tx35h07.x1 NCI-CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2271613 3',  
DEFINITION mRNA sequence.  
ACCESSION AI680969  
VERSION AI680969.1 GI:4891151  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 500)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Dec 20, 1995 this sequence version replaced gi:1135321.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 500.  
Location/Qualifiers  
1..500  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI-CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 114 a 112 c 125 g 145 t  
ORIGIN

Query Match 45.4%; Score 368; DB 50; Length 500;  
Best Local Similarity 99.8%; Pred. No. 2e-189;  
Matches 418; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 230 TCACGCCAGACTCGTGGTGTCTCTGTTCTCAGTCACTACAGGACCTGGGGG 289  
DB 41 TCACGCCAGACTCGTGGTGTCTCTGTTCTCAGTCACTACAGGACCTGGGGG 100  
QY 290 CTGTTGCCACCTCCGCGGGGCGAGGAGTGCCTTAAGTCGAGGACCTCAAAAGTGGAC 349  
DB 101 CTGTTGCCACCTCCGCGGGGCGAGGAGTGCCTTAAGTCGAGGACCTCAAAAGTGGAC 160  
QY 350 AATATATTGTAAAGATCCAAAATAAATGACGCTACGCAAGAACAGTTAACTGTACAA 409  
DB 161 AATATATTGTAAAGATCCAAAATAAATGACGCTACGCAAGAACAGTTAACTGTACAA 220  
QY 410 ACTACAGCTCATGTTTCTCTGTTTCCAGCACCACCAATACTGTAGGATTCAGTG 469  
DB 221 ACTACAGCTCATGTTTCTCTGTTTCCAGCACCACCAATACTGTAGGATTCAGTG 280  
QY 470 GCAATGAAACACATTTTACTGGGAACGAAGTTGTTTTCAGGCCCATATCTTCCGAA 529

Db 281 GCAATGAACACATTTTACTGGACGAGTTGGTTTTTCAAGCCCATATCTTGCGR 340  
QY 530 ATGTAATGGCTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTGGAATGGTGG 589  
Db 341 ATGTAATGGCTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTGGAATGGTGG 400  
QY 590 GAGCAGATCGATTTTACCTTGTATACCCCTGCTTGGGTTGTAAAGTTTTCGACTGTA 648  
Db 401 GAGCAGATCGATTTTACCTTGTATACCCCTGCTTGGGTTGTAAAGTTTTCGACTGTA 459

## RESULT 15

AI903935 AI903935 322 bp mRNA EST 01-DEC-1999  
LOCUS MR-BT038-080199-049 BT038 Homo sapiens cDNA, mRNA sequence.

ACCESSION AI903935

VERSION AI903935.1 GI:6494231

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 322)

HCGP <http://www.ludwig.org.br/ORESTES>.

TITLE The FAPESP/LICR Human Cancer Genome Project

JOURNAL Unpublished (1999)

COMMENT On Dec 20, 1995 this sequence version replaced gi:1134501.

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/seq/gethtml.pl?tl=MR&tl2=MR-BT038-049.html&tl3=080199&tl4=1>)

Seq primer: puc 18 forward.

Location/Qualifiers

1. 322

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BT038"

/sex="female"

/dev\_stage="Adult"

/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 90 a 60 c 68 g 104 t

ORIGIN

Query Match 39.5%; Score 320; DB 62; Length 322;

Best Local Similarity 100.0%; Pred. No. 2.8e-163;

Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GTGGGACAATATTTGTAAGATCCAAAATAAATAGCGCTACGCAAGAACCAAGTTAAC 402

Db 3 GTGGGACAATATTTGTAAGATCCAAAATAAATAGCGCTACGCAAGAACCAAGTTAAC 62

QY 403 GTTACAACTACACAGCTCATGTTTCTCTGTTTCCAGACCCCAACATACTTGTAGGAT 462

Db 63 GTTACAACTACACAGCTCATGTTTCTCTGTTTCCAGACCCCAACATACTTGTAGGAT 122

QY 463 TCCAGTGGCAATGAACACATTTTACTGGGAACGAAGTTGGTTTTTCAAGCCCATATCT 522

Db 123 TCCAGTGGCAATGAACACATTTTACTGGGAACGAAGTTGGTTTTTCAAGCCCATATCT 182  
QY 523 TCCGAAATGTAATGCGTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTTGA 582  
Db 183 TCCGAAATGTAATGCGTATTCCTACAAAGTGGCAGTCGCATGTCTCTTTTCTTGA 242  
QY 583 TGGTTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTGGTTTGGTTTAAAGTTTGC 642  
Db 243 TGGTTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTGGTTTGGTTTAAAGTTTGC 302  
QY 643 ACTGTAGGGTTTGTGGAAT 662  
Db 303 ACTGTAGGGTTTGTGGAAT 322

Search completed: February 10, 2000, 14:59:09

Job time: 5527 sec

"

"

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2000, 17:01:07 ; Search time 11.6 Seconds  
(without alignments)  
297.955 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 1439  
Sequence: 1 MHILKSPNVPRAHGQKNT.....TRLRLSITNETFRKTOLYP 269

Scoring table: BLOSUM62

Searched: 130275 seqs, 12848600 residues

base : Issued Patents\_AA:\*

Word size : 0

Number of hits that pass the threshold : 130275  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pap:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pap:\*  
3: /cgn2\_6/ptodata/2/iaa/PCTUS9.COMB.pap:\*  
4: /cgn2\_6/ptodata/2/iaa/backfiles.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	85	5.9	338	2	US-08-414-657D-42	Sequence 42, Appl
2	85	5.9	338	2	US-08-414-657D-43	Sequence 43, Appl
3	85	5.9	310	2	US-08-414-657D-45	Sequence 45, Appl
4	83	5.8	764	2	US-08-177-109A-2	Sequence 2, Appl
5	83	5.8	764	2	US-08-687-706-2	Sequence 2, Appl
6	80.5	5.6	1237	1	US-08-241-853-2	Sequence 2, Appl
7	80.5	5.6	1237	1	US-08-850-917-2	Sequence 2, Appl
8	80.5	5.6	325	2	US-08-414-657D-2	Sequence 2, Appl
9	80.5	5.6	325	2	US-08-414-657D-41	Sequence 41, Appl
10	80.5	5.6	304	2	US-08-414-657D-44	Sequence 44, Appl
11	79.5	5.5	354	2	US-08-700-013B-9	Sequence 9, Appl
12	79.5	5.5	797	2	US-08-700-013B-21	Sequence 21, Appl
13	79	5.5	797	2	US-08-700-013B-19	Sequence 19, Appl
14	78.5	5.5	338	2	US-08-414-657D-60	Sequence 60, Appl
15	76	5.3	525	1	US-08-356-340-2	Sequence 2, Appl
16	76	5.3	525	2	US-08-786-555-2	Sequence 2, Appl
17	74	5.1	2973	2	US-08-821-355A-7	Sequence 7, Appl
18	74	5.1	1193	2	US-08-400-159-10	Sequence 10, Appl
19	74	5.1	344	2	US-07-857-224B-93	Sequence 93, Appl
20	73.5	5.1	3033	1	US-07-925-695-8	Sequence 8, Appl
21	73.5	5.1	191	2	US-08-290-665A-177	Sequence 177, App
22	73.5	5.1	191	3	PCT-US95-10398-177	Sequence 177, App
23	73	5.1	447	1	US-08-188-281B-20	Sequence 20, Appl
24	73	5.1	381	1	US-08-173-508-12	Sequence 12, Appl
25	73	5.1	627	1	US-08-285-814E-4	Sequence 4, Appl
26	73	5.1	627	1	US-08-291-299-10	Sequence 10, Appl
27	73	5.1	381	2	US-08-265-310-12	Sequence 12, Appl
28	73	5.1	627	3	PCT-US93-01959-4	Sequence 4, Appl
29	73	5.1	447	3	PCT-US94-07280-20	Sequence 20, Appl
30	73	5.1	447	3	PCT-US95-01087-20	Sequence 20, Appl
31	73	5.1	627	3	PCT-US95-10579-10	Sequence 10, Appl
32	73	5.1	328	4	5212074-4	Patent No. 5212074
33	72.5	5.0	3033	1	US-07-925-695-9	Sequence 9, Appl
34	72.5	5.0	615	2	US-08-752-307B-9	Sequence 9, Appl
35	72	5.0	299	2	US-08-872-437-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-414-657D-42  
; Sequence 42, Application US/08414657D

; Patent No. 5861283

; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak

; APPLICANT: Zhukareva, Victoria

; TITLE OF INVENTION: Limbic System-Associated Membrane

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/414,657D

; FILING DATE: 31-MAR-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bloom, Allen

; REGISTRATION NUMBER: 29,135

; REFERENCE/DOCKET NUMBER: 317743-102

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 609-520-3214

; TELEFAX: 609-520-3259

; TELEX:

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 338 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-414-657D-42

Query Match 5.9%; Score 85; DB 2; Length 338;

Best Local Similarity 27.7%; Pred. No. 0.79;

Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

QY 101 SAGGESLKCDELKVG-----QYICKDPKINDATQEPVNCNTNTAHVSCFPAPNITCKDSS 156

Db 230 TTGROASLKCEASAPAPDPFWRDTRINSANGLEIKSTE---GQSSLTAVTNT-EEHY 285

QY 157 GNETHFTGNEVG-----FFKPISCRNVGYKYKVAVALSLFLGWLGDARFYLGYPALG 209



Db 286 GNYTCVAANKLGVNASLVLFRRGSRVING-SISLAVPL-----WL-----LA 328

QY 210 LLKFCVTGFC 219

Db 329 ASLFCLLSKC 338

## RESULT 2

US-08-414-657D-43

; Sequence 43, Application US/08414657D

; Patent No. 5861283

; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak

; APPLICANT: Zhukareva, Victoria

; TITLE OF INVENTION: Limbic System-Associated Membrane

; TITLE OF INVENTION: Protein and DNA

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/414,657D

; FILING DATE: 31-MAR-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bloom, Allen

; REGISTRATION NUMBER: 29,135

; REFERENCE/DOCKET NUMBER: 317743-102

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 609-520-3214

; TELEFAX: 609-520-3259

; TELEX:

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 338 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-414-657D-43

Query Match

Best Local Similarity 5.9%; Score 85; DB 2; Length 338;

Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

QY 101 SAGGEESLKCEDLVG-----QYICKDPKINDATQEPVNCNTNYTAHVSCFPAPNITCKDSS 156

Db 230 TTGRQASLCKEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

QY 157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLEFLGWLGLADRFYLGYPALG 209

Db 286 GNYTCVAANKLGVNASLVLFRRGSRVING-SISLAVPL-----WL-----LA 328

QY 210 LLKFCVTGFC 219

Db 329 ASLFCLLSKC 338

## RESULT 3

US-08-414-657D-45

; Sequence 45, Application US/08414657D

; Patent No. 5861283

; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak

; APPLICANT: Zhukareva, Victoria

; TITLE OF INVENTION: Limbic System-Associated Membrane

; TITLE OF INVENTION: Protein and DNA

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/414,657D

; FILING DATE: 31-MAR-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bloom, Allen

; REGISTRATION NUMBER: 29,135

; REFERENCE/DOCKET NUMBER: 317743-102

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 609-520-3214

; TELEFAX: 609-520-3259

; TELEX:

; INFORMATION FOR SEQ ID NO: 45:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 310 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-414-657D-45

Query Match

Best Local Similarity 5.9%; Score 85; DB 2; Length 310;

Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

QY 101 SAGGEESLKCEDLVG-----QYICKDPKINDATQEPVNCNTNYTAHVSCFPAPNITCKDSS 156

Db 202 TTGRQASLCKEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257

QY 157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLEFLGWLGLADRFYLGYPALG 209

Db 258 GNYTCVAANKLGVNASLVLFRRGSRVING-SISLAVPL-----WL-----LA 300

QY 210 LLKFCVTGFC 219

Db 301 ASLFCLLSKC 310

## RESULT 4

US-08-177-109A-2

; Sequence 2, Application US/08177109A

; Patent No. 5869615

; GENERAL INFORMATION:

; APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby

; TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:



CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1237 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gallus domesticus  
STRAIN: DT40  
US-08-850-917-2

Query Match 5.6%; Score 80.5; DB 1; Length 1237;  
Best Local Similarity 21.6%; Pred. No. 15;  
Matches 48; Conservative 30; Mismatches 75; Indels 69; Gaps 11;  
QY 71 PSAPAVTA-----RLVGLWLFVSVTTGPMGAVATSAGG-----EESLKCE 111  
DB 323 KPAPENLVHPDTRNVTWTM-----KPTGLEKHIDGYTVECNNTSQNNRNETSFTCG 377  
QY 112 DL-----KVGQYICKD---PKINDATQ-----EPVNCNTYTAHVSCFPAPNIT 151  
DB 378 DLEPYSTGVSVPRAFKSKYKKNFEKGVNGSFQTKPAKPNVTDKLTLLADNTVKVA 437  
QY 152 CKDSS--GNETHF-----TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192  
DB 438 CRSQKVYGNETKFLSWSSSSNGENQRKNECF---TVRDLSTLTFTFKISVFNQVY 493  
QY 193 LGWLGADEFYLGYPALGLKFTCTGFCGIGSLIDFILISMQI 234  
DB 494 TGDVSVCEIYTRYNSRALIIFLV--FLIVVTSIALLLVLYKI 533

RESULT 7  
US-08-850-917-2  
Sequence 2, Application US/08850917  
Patent No. 5854045  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanafusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850.917  
FILING DATE: 02-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/241,853  
FILING DATE: 12-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1237 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gallus domesticus  
STRAIN: DT40  
US-08-850-917-2  
Query Match 5.6%; Score 80.5; DB 2; Length 1237;  
Best Local Similarity 21.6%; Pred. No. 15;  
Matches 48; Conservative 30; Mismatches 75; Indels 69; Gaps 11;  
QY 71 PSAPAVTA-----RLVGLWLFVSVTTGPMGAVATSAGG-----EESLKCE 111  
DB 323 KPAPENLVHPDTRNVTWTM-----KPTGLEKHIDGYTVECNNTSQNNRNETSFTCG 377  
QY 112 DL-----KVGQYICKD---PKINDATQ-----EPVNCNTYTAHVSCFPAPNIT 151  
DB 378 DLEPYSTGVSVPRAFKSKYKKNFEKGVNGSFQTKPAKPNVTDKLTLLADNTVKVA 437  
QY 152 CKDSS--GNETHF-----TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192  
DB 438 CRSQKVYGNETKFLSWSSSSNGENQRKNECF---TVRDLSTLTFTFKISVFNQVY 493  
QY 193 LGWLGADEFYLGYPALGLKFTCTGFCGIGSLIDFILISMQI 234  
DB 494 TGDVSVCEIYTRYNSRALIIFLV--FLIVVTSIALLLVLYKI 533

RESULT 8  
US-08-414-657D-2  
Sequence 2, Application US/08414657D  
Patent No. 5861283  
GENERAL INFORMATION:  
APPLICANT: Levitt, Pat  
APPLICANT: Pimenta, Aurea  
APPLICANT: Fischer, Itzhak  
APPLICANT: Zhukareva, Victoria  
TITLE OF INVENTION: Limbic System-Associated Membrane  
TITLE OF INVENTION: Protein and DNA  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 997 Lenox Drive, Building 3, Suite 210  
CITY: Lawrenceville  
STATE: NJ  
COUNTRY: USA  
ZIP: 08543  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/414.657D  
FILING DATE: 31-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bloom, Allen  
REGISTRATION NUMBER: 29,135

Qy	101	SAGGEESLKEDLKVG----	QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS	156
		:   :	:   :      :	:   :
Db	202	TGTRQASLKCEASVAPAPF	EWYRDDTRINSANGLEIKSTE----	QSSSLTVTNVT-EEHY
				257





;; TITLE OF INVENTION: TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
;; STREET: 1180 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: US  
;; ZIP: 10036-8403  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/356,340  
;; FILING DATE: 21-DEC-1994  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US PCT/EP93/01604  
;; FILING DATE: 22-JUN-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: DE P42 20 759.2  
;; FILING DATE: 24-JUN-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meliman, Edward A.  
;; REGISTRATION NUMBER: 24,735  
;; REFERENCE/DOCKET NUMBER: P/951-106  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 382-0700  
;; TELEFAX: (212) 382-0888  
;; TELEX: 236925  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 525 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-356-340-2

Query Match 5.3%; Score 76; DB 1; Length 525;  
Best Local Similarity 28.9%; Pred. No. 12;  
Matches 24; Conservative 6; Mismatches 17; Indels 36; Gaps 3;  
QY 40 LPFSLPL-----GGGGSGGKSVSKMAAAWPGSPSAPAVTARLVGLVWFVSVTTG 93  
: ||: || ||| | : | : |  
P 434 ITFSIPLALASIFASGSGQGLSLGLNLAIIVVPO-----MFVSVTSG 477  
QY 94 PW-----GAVATSA 102  
: ||| : ||  
Db 478 PWDAMFGGGNLPAFVVGAVAATA 500

Search completed: February 10, 2000, 17:01:46  
Job time: 39 sec





Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco  
High quality sequence stop: 459.

#### FEATURES

source

1. 574  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2272257"  
/clone\_lib="NCI-CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"

/note="Organ: lung; Vector: pT730-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP\_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 148 a 118 c 136 g 172 t  
ORIGIN

#### alignment\_scores:

Quality: 1024.00 Length: 191  
Ratio: 5.361 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-060-609-2 x AI680904

Align seg 1/1 to: AI680904 from: 1 to: 574

71 ProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTrpPh 87  
1 CCGTCTGCTCCGAGGCGTGCCTGACTGCCAGACTCGTTGGTGCTCTGGTT 50  
87 eValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyG 104  
51 CGTCTCAGTCACTACAGGACCTGGGGGCTGTGCCACCTCCGCCGGG 100  
104 lYGlUSeuSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCys 120  
101 GCGAGGAGTCGCTTAAGTCCGAGGACCTCAAAGTGGGACATATATTTGT 150  
121 LysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAs 137  
151 AAAGATCCAAAATAAATGACGCTACGCAAGAACACGATTACTGTACAAA 200  
137 nTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysA 154  
201 CTACACAGCTCATGTTCTCTCTTTTCCAGCACCCCAACATCTTGAAGG 250  
154 sPSeuSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePhe 170  
251 ATTCAGTGGCAATGAACACATTTTACTGGGAGACGAGTTGGTTTTTC 300  
171 LysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVa 187  
301 AAGCCCATCTTCCGAAATGTAATGGCTATTCTCTACAAAGTGGCAGT 350  
187 lAlaLeuSerLeuPheLeuGlyTyrLeuGlyAlaAspArgPheTyrLeuG 204  
351 CGCATTTCTCTTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTTG 400  
204 lYTyRProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGly 220  
401 GATACCTGCTTTGGGTTTGTAAAGTTTTCACACTGTAGGGTTTGTGGA 450

1 CGCGCGCGCTGGCGCTGCTGCTCGCTCCGAGGCGCGTGCAGCGCCAG 50  
80 gLeuValGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyA 97  
51 ACTCGTTGGTCTCCTGCTCGTCTCAGTCACTACAGGACCTGGGGGG 100  
97 lAValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113  
101 CTGTTGGCACCTCCCGCGGGGCGAGGAGTCCGCTTAAGTGGGAGACCT 150  
114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130  
151 AAAGTGGGACATATATTTGTAAGATCCAAAATAAATGACGCTACCCA 200  
130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147  
201 AGAACCAAGTAACTGTACAAACTACAGCTCATGTTCTCTGTTTCCAG 250  
147 lAProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163  
251 CACCCACATCAACTGTGAAGATTCCAGTGGCAATGAACACATTTTACT 300  
164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG 180  
301 GGGACGAAGTTGGTTTTTCCAGCCCAATATCTTCCCAAAATGTAATGG 350  
180 yTyRSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTyrLeuG 197  
351 CTATTCCTACAAAGTGCAGCTGCATTCCTCTCTTTCTTCTGGATGGT 400  
197 lYAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLysPhe 213  
401 GAGCAGATCGATTTTACCTTGGATACCTGCTGTTGGGTTTCTGTTAA 450  
214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeu 230  
451 TGCACGTGANGTTTGTGGAATTTGGAGCCTAATGATTTTCATTTAT 500  
230 eSerMetGlnIleValGlyProSerAspGlySerSerTyrIleIleAsp 247  
501 TTCAATGCAGATTGTTGGACCTTCAGATGGAAGTAGITACATATTAG 550  
247 yTyRGlYThrArgLeuThrArgLeuSerIleThrAsnGluThr 261  
551 ACTATGAACACAGACTTACAGACTGAGTATTACTAATGAACA 594

seq name: gb\_est31:AI680904

#### Documentation\_block:

LOCUS AI680904 574 bp mRNA 26-MAY-1999  
DEFINITION tx42f05.x1 NCI-CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2272257 3', similar to WP:02P5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.  
ACCESSION AI680904  
VERSION AI680904.1 GI:4891086  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 574)  
Email: Robert.Strausberg@nih.gov  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Dec 20, 1995 this sequence version replaced gi:1134035.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center

221 ileglsrleuileaspheilleuileSerMetGlnilevalglypr 237  
 451 ATTGGGAGCCTAATGATTCATCTTATTTCAGGAGATGTTGGACC 500  
 237 oSerAspGlySerSertyrilleleAspTyrTyrGlyThrArgLeuThrA 254  
 501 TTCAGATGGAAGTAGTACATATATAGATTACTATGGAACAGACTTACAA 550  
 254 rgLeuSerIleThrAsnGluThr 261  
 551 GACTGAGTATTACTAATCAACA 573

seq\_name: gb\_est28:AI479764

seq\_documentation\_block: 556 bp mRNA EST 14-APR-1999  
 LOCUS AI479764  
 DEFINITION tm69b04.xl NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2163343 3',  
 mRNA sequence.

ACCESSION AI479764  
 SION AI479764.1 GI:4372932  
 WORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 556)  
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BTGAP), Tumor Gene Index

JOURNAL Unpublished (1998)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188819.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/dbbrp/image/image.html

Insert Length: 614 Std Error: 0.00

Seq Primer: -40UP from Gibco  
 High quality sequence stop: 456.

Location/Qualifiers

1..556  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2163343"  
 /clone\_lib="NCI\_CGAP\_Brn25"  
 /tissue\_type="anaplastic oligodendroglioma"  
 /lab\_host="DH10B"

/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 148 a 109 c 130 g 168 t 1 others

ORIGIN

alignment\_scores:  
 Quality: 988.00 Length: 185  
 Ratio: 5.370 Gaps: 0  
 Percent Similarity: 99.459 Percent Identity: 99.459

alignment\_block:

US-09-060-609-2 x AI479764

Align seg 1/1 to: AI479764 from: 1 to: 556

79 AlaArgLeuValGlyValLeuTrrpPheValSerValThrThrGlyProTr 95  
 2 GCCAGACTCGTGTGTCTCTGTTCTCTCAGTCACTACAGGACCCCTG 51  
 95 pGlyAlaValAlaThrSerAlaGlyGlyGluSerLeuLysCysGluA 112  
 52 GGGGCTGTTCACCTCCGCCGGGGGGGAGAGTCGCTTAAGTCGAGG 101  
 112 sPLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAla 128  
 102 ACCTCARAAGTGGACATATATTTGTAAGATCCAAAAATAAATGACGCT 151  
 129 ThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysph 145  
 152 ACGCAAGAACCAGTTAACTGTACAAACTACACAGCTCATGTTCTCTGTT 201  
 145 eProAlaProAsnIleThrCysLysAspSerSerClyAsnGluThrHisP 162  
 202 TCCAGCACCCCAACATACTTGTAAAGATTCCAGTGGCAATGAACACATT 251  
 162 heThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal 178  
 252 TTACTGGGAACCAAGTGGTTTTTTCAGCCCATATCTTGGCAAAATGTA 301  
 179 AsnGlyTyrSertyrLysValAlaValAlaLeuSerLeuPheLeuGlyTr 195  
 302 AATGGCTATTCTTACAAAGTGGCAGTCGCAATTGCTCTCTTTTCTGGATG 351  
 195 pLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuL 212  
 352 GTTCGGAGCAGATGTTACCTTGGATACCTGCTTGGGTTTGTAA 401  
 212 ysPheCysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIle 228  
 402 AGTTTTCACCTGTAGGTTTGTGAATGGAGCTAATGATTTCATT 451  
 229 LeuIleSerMetGlnIleValGlyProSerAspGlySerSertyrIleI 245  
 452 CTTATTTCATGCAATGTTGGACCTTCAGATGGAAGTACATTAT 501  
 245 eAspTyrTyrGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrP 262  
 502 AGATTACTATGGAACCCAGACTTACAAGACTGAGTATTACTAATGAACAT 551  
 262 heArg 263  
 552 TTAGA 556

seq\_name: gb\_est29:AI580361

seq\_documentation\_block:

LOCUS AI580361 537 bp mRNA EST 05-APR-1999  
 DEFINITION tm47h02.xl NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2161299 3',  
 similar to WP:CG2P5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.

ACCESSION AI580361

VERSION AI580361.1 GI:4564737  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 537)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Mar 10, 1998 this sequence version replaced gi:2949103.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
 High quality sequence stop: 492.

## FEATURES

source

1. 537  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2161299"  
 /lab\_host="NCI-CGAP\_Kid11"  
 /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
 124 a 117 c 139 g 156 t 1 others

BASE COUNT  
 ORIGIN

## alignment\_scores:

Quality: 957.00 Length: 179  
 Ratio: 5.376 Gaps: 0  
 Percent Similarity: 99.441 Percent Identity: 98.883

## alignment\_block:

US-09-060-609-2 x AI580361 ..

Align seg 1/1 to: AI580361 from: 1 to: 537

64 AlaAlaIaIaTrpProSerGlyProSerAlaProGluAlaValThrAlaAr 80  
 1 GCGGCGCGCTGGCGTGTGTCGTCTCAGTCACTACAGGACCTGGGGG 50  
 80 GLeuValGlyValLeuTrpPheValSerValThrGlyProTrpGlyA 97  
 51 ACTCGTGTGTCCTGTGTCGTCTCAGTCACTACAGGACCTGGGGG 100  
 97 lValAlaThrSerAlaGlyGlyGluGluSerLeuLysCysGluAspLeu 113  
 101 CTGTGTCACACCTCCGCGGGGCGAGGAGTCGTTAAAGTCGAGGACCTC 150  
 114 LysValGlyGlnTyrlleCysLysAspProLysIleAsnAspAlaThrCl 130  
 151 AAAGTGGGCAATATATTGTAAGATCCAAAATAAATGACGCTACGCA 200  
 130 nGluProValAsnCysThrAsnTyThrAlaHisValSerCysPheProA 147  
 201 AGAACCACTTAAGTACAACTACACAGCTCATGTTCTCTGTTTCCAG 250  
 147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163  
 251 CACCAACACATACTTGAAGGATTCAGTGGCAATGAACACATTTTACT 300  
 164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnGl 180  
 301 GGGAAACGAGTGGTGGTTTTTCAAGCCCATATCTTGCAGAAATGTAATGG 350

180 YTrSerTyrlsValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197  
 351 CATTCTCTCAAGTGGCGATGTCATTCTTTCTGGATGGTGG 400  
 197 lYAlaAspArgPheTyrlsValAlaLeuGlyProAlaLeuGlyLeuLysPhe 213  
 401 GAGCAGATCGATTACCTGGATACCTGCTNTGGTGTGTTAAAGTTT 450  
 214 CysThrValGlyPheCysGlyLeuGlySerLeuIleAspPheIleLeu 230  
 451 TGCACGTAGGGTTTGTGGAATGGAGCTAATGATTCTTCATTCTAT 500  
 230 eSerMetGlnIleValGlyProSerAspGlySerSer 242  
 501 TTCAATGCAGATTGTTGGACCTTCAGATGGGAAGAGT 537  
 seq\_name: gb\_est31:AI674462

seq\_documentation\_block:  
 LOCUS AI674462 530 bp mRNA EST 19-MAY-1999  
 DEFINITION wc44e01.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2321496 3', similar to WP:CO2F5.3 CE00039 GTP-BINDING PROTEIN ; mRNA sequence.  
 ACCESSION AI674462  
 VERSION AI674462.1 GI:4874942  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 530)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948555.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
 High quality sequence stop: 459.

## FEATURES

source

1. 530  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2321496"  
 /clone\_lib="NCI-CGAP\_Pr28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"

/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP\_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."  
 123 a 116 c 133 g 157 t 1 Others

BASE COUNT  
 ORIGIN

## alignment\_scores:

Quality: 947.00

Length: 176

```
Ratio: 5.411 Gaps: 0
Percent Similarity: 99.432 Percent Identity: 99.432
alignment_block:
US-09-060-609-2 x AI674462 ..
Align seg 1/1 to: AI674462 from: 1 to: 530

66 AlaTrpProSerGlyProSerAlaProGluAlaValThrAlaArgLeuVal 82
|||||
3 GCCTGGCGTCGTGGTCGTCTGCTCCGGAGCGGTACGGCCAGACTCGT 52
82 lGlyValLeuTrpPheValSerValThrThrGlyProTrpGlyAlaVal 99
|||||
53 TGTGTCTGTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 102
99 laThrSerAlaGlyGlyGluSerLeuLysCysGluAspLeuLysVal 115
|||||
103 CCACCTCCCGCGGGCGAGGAGTCTTAAGTCCGAGGACCTCAAGTG 152
116 GlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluPr 132
|||||
153 GGACATATATTGTAAAGATCCAAAAATAAATGACGCTACGCAAGACC 202
132 oValAsnCysThrAsnTyrThrAlaHisValSerCysPheProAlaProA 149
|||||
203 AGTTAACTGTACAACTACACAGCTCATGTTCTCTGTTTCCAGCACCCA 252
149 snIleThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsn 165
|||||
253 ACATAACTTGTAGGATCCAGTGGCAATGAAACACATTTTACTGGGAAC 302
166 GluValGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSe 182
|||||
303 GAAGTTGGTTTTTCAAGCCCATATCTTCCGAAATGTAATGGCTATTC 352
182 rTyLysValAlaValAlaLeuSerLeuPheLeuGlyTyrTrpLeuGlyAla 199
|||||
353 CTACAAAGTGGCAGTGCATCTCTCTTTCTTGGATGGTGGGAGCAG 402
199 spArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLysPheCysThr 215
|||||
403 ATCGATTTTACCTTGGATACCTCGCTTGGTTTGTAAAGNTTTCACCT 452
216 ValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeuIleSerMe 232
|||||
453 GTAGGGTTTTTGGGAATGGGAGCCCAATGTGATTTTCATCTTTATTCAT 502
232 tGlnIleValGlyProSerAspGlySer 241
|||||
503 GCAGATTGTGGACCTTCAGATGGAGT 530

seq_name: gb_est19:AA772225

seq_documentation_block:
LOCUS AA772225 548 bp mRNA EST 31-DEC-1998
DEFINITION ai41c01.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
1359552.3 similar to WP:002F5.3 CE00039 GTP-BINDING PROTEIN ;
mRNA sequence.
ACCESSION AA772225
VERSION AA772225.1 GI:2824008
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 548)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:811076.
Contact: Robert Strausberg, Ph.D.
```

Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyt not found  
Insert Length: 978 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 459.

FEATURES  
Location/Qualifiers  
1..548  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="1359552"  
/clone\_lib="Soares\_parathyroid\_tumor\_NbHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer  
15'-  
TGTTACCAATCTGAAGTGGAGCGCGCACCAATTTTTTTTTTTTTTTTTTTT  
T-3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 146 a 106 c 126 g 169 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 924.50 Length: 184  
Ratio: 5.136 Gaps: 3  
Percent Similarity: 97.826 Percent Identity: 96.196

alignment\_block:  
US-09-060-609-2 x AA772225 ..  
Align seg 1/1 to: AA772225 from: 1 to: 548

81 LeuValGlyValLeuTrpPheVal.SerValThrThrGlyProTrpGlyA 97  
|||||  
2 CTGCTGGTGTCTGTGGTTCGTCTGTCAGTCACAGACCTCGGGGG. 50  
97 laVal.AlalaThrSerAlaGlyGlySerLeuLysCysGluAspLe 113  
|||  
51 ..CTGTTGCACCTCCGCGGGGGCANGAGTGCCTTAAGTCGAGGACCT 98  
113 uLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130  
|||||  
99 CAAAGTGGGCAATATATTGTAAGATCCAAAAATAAATGACGCTACGC 148  
130 lGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146  
|||||  
149 AAGAACAGTAACTGTACAACTACACAGCTCATGTTCTCTGTTTCCA 198  
147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163  
|||||  
199 GCACCCCAACATACTTGTAAAGATTCAGTGCGCAATGAACACATTTTAC 248

163 rGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG 180  
 |||||  
 249 TGGAGCAAGATTGGTTTTCAGAGCCATATCTGCGGAAATGTAATG 298  
 |||||  
 180 lYTySerTyLysValAlaValAlaLeuSerLeuPheLeuGlyTriLeu 196  
 |||||  
 299 GCTATCTACAAAGTGGCAGTCGATGTCTCTTTCTTGGATGGTTG 348  
 |||||  
 197 GlyAlaAspArgPheTyLeuGlyTyProAlaLeuGlyLeuLeuLysPh 213  
 |||||  
 349 GGAGCAGATCGATTTTACCTTGGATACCTGCTTTGGGTTTGTAAAGTT 398  
 |||||  
 213 eCysThrValGlyPheCysGlyIleGlySerLeuLeuAspPheIleu 230  
 |||||  
 399 TTGCACGTAGGGTTTGTGGAATTGGAGGCTAATTGATTTCATTCTTA 448  
 |||||  
 230 leSerMetGlnIleValGlyProSerAspGlySerSerTyIleIleAsp 246  
 |||||  
 449 TTTCAATGCAGATTGTGGACCTTCAGATGGAAGTAGTTACATTATAGAT 498  
 |||||  
 247 TyTyGlyThrArgLeuThrArgLeuSerIleThrAsnGluThrPhe 262  
 |||||  
 499 TACTATGGAACACAGATACAAAGCTGAGTATTACTAATGAACATTT 546  
 |||||

seq\_name: gb\_est31:AI682204

seq\_documentation\_block: 507 bp mRNA EST 26-MAY-1999  
 LOCUS AI682204  
 DEFINITION wa7lb06.xl Soares\_NFL.T.GBC.S1 Homo sapiens cDNA clone  
 IMAGE:2301587 3', mRNA sequence.  
 ACCESSION AI682204  
 VERSION AI682204.1 GI:4892386  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 507)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1136077.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 469.

## FEATURES

source  
 1..507  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2301587"  
 /clone\_lib="Soares\_NFL.T.GBC.S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-CGAP.GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 120 a 109 c 125 g 152 t 1 others

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 905.00 Length: 169  
 Ratio: 5.387 Gaps: 0  
 Percent Similarity: 99.408 Percent Identity: 99.408  
 alignment\_block:  
 US-09-060-609-2 x AI682204  
 Align seg 1/1 to: AI682204 from: 1 to: 507  
 70 GlyProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTr 86  
 |||||  
 1 GGTCCGCTGCTCTCCGAGGCCGTGACGCCAGACTCGTTGGTGTCTCTG 50  
 |||||  
 86 pPheValSerValThrGlyProTriPglyAlaValAlaIleThrSerAla 103  
 |||||  
 51 GTTCGCTCAGTCACTACAGGACCTGGGGGCTGTGCCACCTCCGCCG 100  
 |||||  
 103 lYgLYGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyIle 119  
 |||||  
 101 GGGCGGAGGAGTCGCTTAAGTCGAGGACCTCAAAGTGGGCAATATAT 150  
 |||||  
 120 CysLysAspProLysIleAsnAspAlaThrGlnGluProValAsnCys 136  
 |||||  
 151 TGTAAAGATCCAAATAATGACCTACGCAAGAACCACTTAACGTGAC 200  
 |||||  
 136 rAsnTyThrAlaHisValSerCysPheProAlaProAsnIleThrCysL 153  
 |||||  
 201 AAACCTACACAGCTCATGTTCCTGTTTCCAGCACCCCAACATACTGTA 250  
 |||||  
 153 ysAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhe 169  
 |||||  
 251 AGGATTCCAGTGGCAATGAACACACATTTTACTGGGAACGAAGTTGTTT 300  
 |||||  
 170 PheLysProIleSerCysArgAsnValAsnGlyTySerTyLysValAl 186  
 |||||  
 301 TTCAGGCCATATCTTGCAGAAATGTAATGGCTATTCTACAAAGTGGC 350  
 |||||  
 186 aValAlaLeuSerLeuPheLeuGlyTyTriPleuGlyAlaAspArgPheTy 203  
 |||||  
 351 AGTCGATGCTCTCTTTTCTTGGATGGTTGGAGCAGATCGATTTTACC 400  
 |||||  
 203 euGlyTyProAlaLeuGlyLeuLysPheCysThrValGlyPheCys 219  
 |||||  
 401 TTGGATACCTGCTCTTTTGTAAAGCTTTTGCACCTGTANGGTTTGT 450  
 |||||  
 220 GlyIleGlySerLeuIleAspPheIleLeuIleSerMetGlnIleValG 236  
 |||||  
 451 GGAATTGGAGCCTAATTGATTCTTCTTATTTCATGCAGATTGTTGG 500  
 |||||  
 236 yProSer 238  
 |||||  
 501 ACCTTCA 507  
 seq\_name: gb\_est26:AI349520  
 seq\_documentation\_block: 498 bp mRNA EST 03-FEB-1999  
 LOCUS AI349520  
 DEFINITION qp72905.xl Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
 IMAGE:1928600 3', mRNA sequence.  
 ACCESSION AI349520  
 VERSION AI349520.1 GI:4086726  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 498)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.





|||||  
448 TTGCACTAGAGCTTTGTGGAAATGGGAGCCTAATGTGATTCATCTT 497

230 IleSerMet 232

|||||

498 ATTTCAATG 506

seq\_name: gb\_est24:AI168073

seq\_documentation\_block: 471 bp mRNA EST 05-OCT-1998  
LOCUS AI168073  
DEFINITION o289c02.x1 Soares\_senescent\_fibroblasts\_NbHSF Homo sapiens cDNA  
clone IMAGE:1682498 3', mRNA sequence.

ACCESSION AI168073

VERSION AI168073.1 GI:3701243

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 471)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2282128.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: -40ml3 fwd. ET from Amersham.

#### FEATURES

source

1..471  
Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1682498"

/tissue\_type="senescent\_fibroblasts\_NbHSF"

/lab\_host="DH10B (ampicillin resistant)"

/note="Vector: pT73D (Pharmacia) with a modified

polylinker V-type; phagemid; Site\_1: Not I; Site\_2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5,

TGTTACCAATCTGAAGTGGGAGCGCCGCTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo."

111 a 110 c 120 g 130 t

BASE COUNT

ORIGIN

#### alignment\_scores:

Quality: 805.50 Length: 182

Ratio: 5.197 Gaps: 2

Percent Similarity: 85.165 Percent Identity: 85.165

#### alignment\_block:

US-09-060-609-2 x AI168073 ..

Align seg 1/1 to: AI168073 from: 1 to: 471

64 AlaAlaIatPProSerGlyProSerAlaProGluAlaValThraAlaR 80

|||||

1 GCGCGCGCTGGCGCTGTGCTGCTCGGAGCGCGTACGCGCAG 50

|||||

80 gLeuValGlyValLeuTrpPheValSerValThrGlyProTrpGlyA 97

|||||

51 ACTCGTGTGTCCTGTCGTCTCTCAGTCACTACAGGACCCCTGGGGG 100

97 laValAlaIatPProSerAlaGlyGlyGluSerLeuLysCysGluAspLeu 113

|||||

101 CTGTTGCCACCTCCGCGGGGCGAGAGTCGCTTAAGTCGAGAGCCTC 150

|||||

114 LysValGlyGlnTrpIleCysLysAspProLysIleAsnAspAlaThrG1 130

|||||

151 AAAGTGGGACAAATATTTGTAAGATCCAAAATAAATGACGCTACGCA 200

|||||

130 nGluProValAsnCysThrAsnTrpThrAlaHisValSerCysPheProA 147

|||||

201 AGAACCAAGTTAACTGACAAACTACAGCTCATGTTCTCTGTTTCCAG 250

|||||

147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163

|||||

251 CACCCAACATAACTTGTGAAGATTCACAGTGGCAATGAACACACATTTTACT 300

|||||

164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG1 180

|||||

301 GGGAAACGAAGTTGGTTTTTCAAGCCCATATCTTCCGAAATGTAATGG 350

|||||

180 yTrpSerTrpLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197

|||||

351 CTATTCTCAAAAGTGGCAGTCGCATGTCTCTTTTCTTGCGATGGTGG 400

|||||

197 lyAlaAspArgPheTrpLeuGlyTrpProAlaLeuGlyLeuLysPhe 213

|||||

401 GAGCAGATCGATTTTACCTTGGATACCTGCT..... 432

|||||

214 CysThrValGlyPheCysGlyIleGlySerLeuIleAspPheIleLeu11 230

|||||

432 ..... 432

230 eSerMetGln.IleValGlyProSerAspGlySerSerTyrIle 244

|||||

433 .....TTGGATGTGTGGACCTTCAGATGAAGTAGTTACATT 469

|||||

seq\_name: gb\_est25:AI299154

seq\_documentation\_block:

LOCUS AI299154 439 bp mRNA EST 29-JAN-1999

DEFINITION qm96f11.x1 NCI-CGAP\_Du5 Homo sapiens cDNA clone IMAGE:1896621 3',

mRNA sequence.

ACCESSION AI299154

VERSION AI299154.1 GI:39588808

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

Tumor Gene Index, National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/db/rrp/image/image.html](http://www-bio.llnl.gov/db/rrp/image/image.html)

Insert Length: 966 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 419.

Location/Qualifiers

1..439

/organism="Homo sapiens"

/db\_xref="taxon:9606"

source



/clone="IMAGE:1896621"  
 /clone\_lib="NCI\_CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT73b-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 105 a 100 c 110 g 124 t  
 ORIGIN

alignment\_scores:  
 Quality: 796.00 Length: 146  
 Ratio: 5.452 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-060-609-2 x A1299154

Align seg 1/1 to: A1299154 from: 1 to: 439

```

70 GlyProSerAlaProGluAlaValThrAlaArgLeuValGlyValLeuTr 86
|||||
1 GGTCCGCTGCTCGGAGCGCGTACGCGCAGACTCGTTGGTGTCTGTG 50
|||||
86 pPheValSerValThrGlyProTrpGlyAlaValAlaThrSerAlaG 103
|||||
51 GTTCGTCTCAGTACTACAGGACCTCGGGGGCTGTGCCACTCGCCG 100
|||||
103 lyGlyGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIle 119
|||||
101 GGGCGAGGAGTCTTAAGTCGAGGACCTCAAGTGGGACAATATATT 150
|||||
120 CysLysAspProLysIleAsnAspAlaThrGlnGluProValAsnCysTh 136
|||||
151 TGTAAAGATCCAAAATAATACGCTACGCAAGAACCACTTAACGTGAC 200
|||||
136 rAsnTyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysL 153
|||||
201 AAATACACAGCTCATGTTCTGTTTCCAGCACCAACATACTTGTGA 250
|||||
153 yAspSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhe 169
|||||
251 AGGATCCAGTGGCAATGAACACATTTACTGGGACCAAGTTGGTTT 300
|||||
170 PheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAl 186
|||||
301 TTCAAGCCCATATCTTGGCGAAATGAATGGCTATTCTCACAAAGTGGC 350
|||||
186 aValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrL 203
|||||
351 AGTCGATGTCCTCTTTCTTGGATGGTTGGGACGACATCGATTTTACC 400
|||||
203 euGlyTyrProAlaLeuGlyLeuLeuLysPheCysThr 215
|||||
401 TTGGATACCTGCTTGGTGTGTTAAAGTTTGTCACT 438
|||||

```

seq\_name: gb\_est12:AA306979

seq\_documentation\_block:  
 LOCUS AA306979 475 bp mRNA EST 18-APR-1997  
 DEFINITION EST178050 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end,  
 mRNA sequence.  
 ACCESSION AA306979  
 VERSION AA306979.1 GI:1959469  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

#### REFERENCE AUTHORS

1 (Bases 1 to 475)  
 Adams M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
 Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
 White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,  
 Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
 Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.,  
 Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
 Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
 Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,  
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
 Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
 He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
 Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H.,  
 Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,  
 Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
 Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)

#### TITLE JOURNAL MEDLINE COMMENT

On Sep 12, 1996 this sequence version replaced gi:1288147.

Other ESTs: THC195202

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

#### FEATURES

Location/Qualifiers

1..475

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):108677"

/db\_xref="taxon:9606"

/clone\_lib="Colon carcinoma (HCC) cell line"

/tissue\_type="colon"

/cell\_type="KM12C"

/cell\_line="KM12C(HCC)-parental human colon

carcinoma/Dukes B2"

/note="Organ: colon; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI"

BASE COUNT 108 a 107 c 119 g 135 t 6 others

#### ORIGIN

#### alignment\_scores:

Quality: 791.00 Length: 159

Ratio: 5.136 Gaps: 1

Percent Similarity: 96.855 Percent Identity: 96.226

#### alignment\_block:

US-09-060-609-2 x AA306979

Align seg 1/1 to: AA306979 from: 1 to: 475

```

64 AlaAlaLaTtpProSerGlyProSerAlaProGluAlaValThrAlaAr 80
|||||
2 CGCGCGCGCTGCGCTTCGTCGTCGCGAGCGCGTACGCGCAG 50
|||||
80 gluValGlyValLeuThrPheValSerValThrThrGlyProTrpGlyA 97
|||||
51 ACTCGTGTGTCTGTCGTGGTGTCTCAGTCACTACAGGACCTGGGGGG 100
|||||
97 laValAlaThrSerAlaGlyGlyGluSerLeuLysCysGluAspLeu 113
|||||

```

```
101 CTGTTGCCACCTCCGCGGGGGGAGAGTCGCTTAAGTCGGAGGACCTC 150
114 LysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrG1 130
151 AAAGTGGACAATATATTNTAAAGATCCAAATAATAATNACGCTACGCA 200
130 nGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPheProA 147
201 AGAACCAAGTTAACTGACAAACTACACAGCTCATGTTCTCTGTTTCCAG 250
147 laProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheThr 163
251 CACCCACATRACTTGTAGGATTCAGTGGCAATGAACACATTTTACT 300
164 GlyAsnGluValGlyPhePheLysProIleSerCysArgAsnValAsnG1 180
301 GGGACGAANTNGTTTTTCAAGCCCATATCTTCGCGAAATGTAAATGG 350
180 yTyrSerTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuG 197
351 CTATTCTCAAAAGTGGCAGTCGATGTCTCTTTTCTTGGATGGTTGG 400
197 LysAlaAspArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLys.Ph 213
401 GAGCAGATCATTTTACCTTGGATACCTCTTGGGTTTGTAAAGTTT 450
213 eCysThrValGlyPheCysGlyIle 221
451 TTGCAGTGTAGGGTTTGTGGAAATT 475

seq_name: gb_est22:AI038331
seq_documentation_block:
LOCUS AI038331 448 bp mRNA EST 24-SEP-1998
DEFINITION ox84a06.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
clone IMAGE:1662994 3', mRNA sequence.
ACCESSION AI038331
VERSION AI038331.1 GI:3277525
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1099 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 432.
location/Qualifiers
1..448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1662994"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V_PYE; phagemid: Site.1: Not I; Site.2: Eco
RI; lstr strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTCACCAATCTGAAGTGGAGCGGCCGACATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
```

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 118 a 92 c 105 g 133 t  
ORIGIN

alignment\_scores:  
Quality: 779.00 Length: 146  
Ratio: 5.372 Gaps: 0  
Percent Similarity: 99.315 Percent Identity: 98.630

alignment\_block:  
US-09-060-609-2 x AI038331 ..  
Align seg 1/1 to: AI038331 from: 1 to: 448

88 ValSerValThrThrGlyProTrpGlyAlaValAlaThrSerAlaGlyG1 104  
|||||  
5 GTCTCAGTCACCTACAGGACCTGGGGGCTGTGCCACCTCCGCGGGG 54  
104 YGluGluSerLeuLysCysGluAspLeuLysValGlyGlnTyrIleCysL 121  
|||||  
55 CGAGGAGTCGCTTAAGTCGGAGGACCTCAAAGTGGGCAATATATTTGTA 104  
121 YSAspProLysIleAsnAspAlaThrGlnGluProValAsnCysThrAsn 137  
|||||  
105 AAGATCCAAAATAAATGACCTACGCAAGAACCACTTAAGTGTACAAAC 154  
138 TyrThrAlaHisValSerCysPheProAlaProAsnIleThrCysLysAs 154  
|||||  
155 TACACAGCTCATGTTCTCTGTTTCCAGCACCCCAACATACTTGTGAAGA 204  
154 pSerSerGlyAsnGluThrHisPheThrGlyAsnGluValGlyPhePheL 171  
|||||  
205 TTCAGTGGCAATGAACACATTTTACTGGGAGCAAGATTTGGTTTTTCA 254  
171 YSProIleSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaVal 187  
|||||  
255 AGCCATATCTTGGCGAATGTAATGGCTATTCTCAAAAGTGGCAGTC 304  
188 AlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuG1 204  
305 GCATTGCTCTCTTTCTTGGATGCTGGGACGACATGATTTTACCTTGG 354  
204 yTyrProAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyI 221  
|||||  
355 ATACCTGCTTTGGTTTGTAAAGTTTTCACCTGTATGTTATGTGGA 404  
221 leGlySerLeuIleAspPheIleLeuIleSerMetGln 233  
|||||  
405 TTGGAGCCCAATGATTTCATCTTATTTCATCAATGCAG 442

seq\_name: gb\_est8:AA017855

seq\_documentation\_block: 488 bp mRNA EST 21-JAN-1997  
LOCUS AA017855  
DEFINITION mh47c09.r1 Soares mouse placenta 4NbMp13.5 14.5 Mus musculus cDNA  
clone IMAGE:445648 5' similar to WP:CO2F5.3 CE00039 GTP-BINDING  
PROTEIN ;, mRNA sequence.

ACCESSION AA017855  
VERSION AA017855.1 GI:1481174  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

# TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Apr 14, 1993 this sequence version replaced gi:785679.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:270984

Possible reversed clone: similarity on wrong strand  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 472.

## FEATURES

### source

1. .488  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:445648"  
/clone\_lib="Soares mouse placenta 4NDMP13.5 14.5"  
/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="PH10B"  
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'  
TGTACCACATCTGAATGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Soares and M.Fatima Bonaldo."

BASE COUNT 150 a 101 c 87 g 150 t  
ORIGIN

## alignment\_scores:

Quality: 777.00 Length: 154  
Ratio: 5.112 Gaps: 0  
Percent Similarity: 98.701 Percent Identity: 95.455

## alignment\_block:

09-060-609-2 x AA017855/rev ..

Align seg 1/1 to reverse of: AA017855 from: 1 to: 488

116 GlyGlnTyrIleCysLysAspProLysIleAsnAspAlaThrGlnGluPr 132  
|||||  
471 GGACAAATATTTTGAAGAACCACCAATAAATGATCG.ACGCAAGAACCC 423  
132 oValAsnCysThrAsnTyrThrAlaHisValSerCysPheProAlaProA 149  
|||||  
422 AGTTAATGTACAACTACACAGCTCATGTTCAATGTTTCCAGCACCCA 373  
149 snlleThrCysLysAspSerSerGlyAsnGluThrHisPheThrGlyAsn 165  
:|||||  
372 AAATAACTTGTAGGATTTGAGTGTAATGAACACACATTTTACTGGAAGT 323  
166 GluValGlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSe 182  
|||||  
322 GAAGTCGGTTTTCACAGCCCATATCTGCCAAATGTGAATGGCTATTC 273  
182 rTyrLysValAlaValAlaLeuSerLeuPheLeuGlyTrpLeuGlyAlaA 199  
|||||  
272 GTACAAAGTGCAGCTGCATATCTCTCTTTTGGGATGGCTGGAGCAG 223  
199 spArgPheTyrLeuGlyTyrProAlaLeuGlyLeuLeuLysPheCysThr 215  
|||||

222 ATCGATTTTACCTCGGATATCCTGCCTTAGGCTTGTTAAATTTTCACC 173  
216 ValGlyPheCysGlyIleCysLeuIleAspPheIleLeuIleSerMe 232  
|||||  
172 GTAGGATTTTGGGAATTTGGAGCCTAATTGATTCATCTTATTTCAAT 123  
232 tGlnIleValGlyProSerAspGlySerSerTyrIleIleAspTyrTyrG 249  
|||||  
122 GCAGATTTGTTGGACCTTCAGATGGAAGTAGTTACATTATAGACTATTATG 73  
249 lyThrArgLeuThrArgLeuSerIleThrAsnGluThrPheArgLysThr 265  
|||||  
72 GAACAGGCTTACAGACTCAGCATTTACTACTATGAAACATTTAGAAAACC 23  
266 GlnLeuTyrPro 269  
|||||  
22 CAGCTGTACCCA 11

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2000, 04:15:42 ; Search time 13.34 Seconds  
(without alignments)  
477.628 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 1439  
Sequence: 1 MHILKSPNVIPRAHQKNT.....TRLRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36:\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1439	100.0	269	1 W94291	Human beta-amyloid
2	748	52.0	139	1 Y12358	Human 5' EST seque
3	667.5	46.4	148	1 Y12426	Human 5' EST seque
4	88	6.1	764	1 W94360	Human Factor B ana
5	86	6.0	764	1 W94365	Human Factor B ana
6	85	6.0	764	1 W94372	Human Factor B ana
7	85	5.9	338	1 W05153	Rat LAMP residues
8	85	5.9	338	1 W05154	Rat LAMP residues
9	85	5.9	310	1 W05156	Rat mature LAMP. D
10	85	5.9	764	1 W94361	Human Factor B ana
11	84	5.8	764	1 W94366	Human Factor B ana
12	84	5.8	764	1 W94373	Human Factor B ana
13	83.5	5.8	763	1 W94362	Human Factor B ana
14	83	5.8	764	1 W94383	Human Factor B ana
15	83	5.8	764	1 W94384	Human Factor B ana
16	83	5.8	764	1 W94385	Human Factor B ana
17	83	5.8	764	1 W94386	Human Factor B ana
18	83	5.8	764	1 W94387	Human Factor B ana
19	83	5.8	764	1 W94388	Human Factor B ana
20	83	5.8	764	1 W94389	Human Factor B ana
21	83	5.8	762	1 W94390	Human Factor B ana
22	83	5.8	764	1 W94352	Human Factor B ana
23	83	5.8	764	1 W94353	Human Factor B ana
24	83	5.8	764	1 W94354	Human Factor B ana
25	83	5.8	764	1 W94355	Human Factor B ana
26	83	5.8	764	1 W94356	Human Factor B ana
27	83	5.8	764	1 W94359	Human Factor B ana
28	83	5.8	764	1 W94377	Human Factor B ana
29	83	5.8	764	1 W94378	Human Factor B ana
30	83	5.8	764	1 W94379	Human Factor B ana
31	83	5.8	764	1 W94380	Human Factor B ana
32	83	5.8	764	1 W94381	Human Factor B ana
33	83	5.8	763	1 W94382	Human Factor B ana
34	82	5.7	409	1 W43449	Tobacco laccase cl
35	82	5.7	764	1 W94364	Human Factor B ana
36	82	5.7	764	1 W94376	Human Factor B ana
37	81	5.6	797	1 W56378	Human Glyt-2 trans
38	80.5	5.6	325	1 W05152	Human LAMP residue
39	80.5	5.6	304	1 W05155	Human mature LAMP.

40	80.5	5.6	1237	1 W44729	Chicken protein ty
41	80.5	5.6	456	1 W69555	Human equilibrativ
42	80.5	5.6	373	1 W64550	Human fibrosarcoma
43	80.5	5.6	1237	1 W89347	Chicken transmembr
44	80.5	5.6	763	1 W94367	Human Factor B ana
45	80.5	5.6	763	1 W94358	Human Factor B ana

ALIGNMENTS

RESULT 1

ID W94291 standard; Protein; 269 AA.  
AC W94291:  
DE Human beta-amyloid peptide-binding protein (BBP).  
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;  
human; Alzheimer's disease.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 58..269  
FT /note= "specifically claimed fragment having  
beta-amyloid peptide binding activity"  
PN W09846636-A2.  
PD 22-OCT-1998.  
PF 14-APR-1998; U07462.  
PR 16-APR-1997; US-064583.  
PA (AMHP ) AMERICAN HOME PROD CORP.  
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,  
Walker SG;  
DR WPI: 99-080736/07.  
DR N-PSDB: X05735.  
PT Polynucleotide encoding beta-amyloid peptide binding protein - used  
to identify inhibitors of beta-amyloid peptide for treating  
Alzheimer's disease  
PS Claim 7; Pages 43-44; 59pp; English.  
CC The present sequence represents a beta-amyloid peptide binding protein  
(BBP). The polynucleotide comprising the entire BBP nucleotide sequence  
of clone BBP1-fl is deposited under the accession number ATCC 98617. The  
polynucleotide comprising a fragment of BBP (nucleotides 202-807 of the  
full length BBP) of clone PEK196 is deposited as ATCC 98399. Host cells  
transformed with a vector comprising the BBP nucleic acid are used for  
the recombinant production of the protein. The protein can be used in a  
method for diagnosing a disease characterised by aberrant expression of  
human beta-amyloid protein (BAP). The protein can also be used in a  
method for screening for compounds which regulate expression of a BAP  
binding protein. The proteins, antibodies and identified compounds can be  
used in the treatment or prevention of Alzheimer's disease.  
SQ Sequence 269 AA;

Query Match	100.0%	Score 1439;	DB 1;	Length 269;
Best Local Similarity	100.0%;	Pred. No. 5.1e-137;		
Matches 269;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps	0;			
QY	1	MHILKSPNVIPRAHQKNTNRDGTGLYPMRGPFKNLALLPFLPLGGGGSGGSEKVS	60	
DB	1	MHILKSPNVIPRAHQKNTNRDGTGLYPMRGPFKNLALLPFLPLGGGGSGGSEKVS	60	
QY	61	SKMAAWPSPSAPEAVTARLYGLVLFVSVTTGTPWCAVATSAAGESLKCEDLVKGYIC	120	
DB	61	SKMAAWPSPSAPEAVTARLYGLVLFVSVTTGTPWCAVATSAAGESLKCEDLVKGYIC	120	
QY	121	KDPKINDATQEPVNCNTYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFKPISCRNVNG	180	
DB	121	KDPKINDATQEPVNCNTYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFKPISCRNVNG	180	
QY	181	YSYKVAVALSLFLGWLGADEFYIGYPALGLLKFKCTGFCGIGSLIDFILISMQIVGSDG	240	
DB	181	YSYKVAVALSLFLGWLGADEFYIGYPALGLLKFKCTGFCGIGSLIDFILISMQIVGSDG	240	
QY	241	SSYIIDYGTTRULSITNETFRKTQLYP	269	

```
|||||
Db 241 SSIIIDYGRRLRLSLTNETFRKQLYP 269
|||||
RESULT 2
YL2358
ID YL2358 standard; Protein: 139 AA.
AC YL2358;
DE 17-JUN-1999 (first entry)
KW Human 5' EST secreted protein SEQ ID NO:389.
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IBL222.
PR 01-AUG-1997; US-905135.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR N-PSDB: X41191.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 27; Page 714-715; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC chemokine activity, haemostatic and thrombolytic activity, chemotactic/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 139 AA.

Query Match 52.0%; Score 748; DB 1; Length 139;
Best Local Similarity 99.3%; Pred. No. 6.5e-68;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 63 MAAAWSPGSAPAVTARLVGLVFWFVSVTTGPGAVATSAGGESLKCEDLKVGQYICKD 122
Db 1 MAAAWSPGSAPAVTARLVGLVFWFVSVTTGPGAVATSAGGESLKCEDLKVGQYICKD 60
QY 123 PKINDATQEPVNCNTNTAHVSCFPAPNITCKDSGNETHTFTGNEVGFPPISCRNVNGYS 182
Db 61 PKINDATQEPVNCNTNTAHVSCFPAPNITCKDSGNETHTFTGNEVGFPPISCRNVNGYS 120
QY 183 YKVAVALSLFLWLGADRF 201
Db 121 YKVAVALSLFLWLGADRF 139

RESULT 3
YL2426
ID YL2426 standard; Protein: 148 AA.
AC YL2426;
DE 17-JUN-1999 (first entry)
KW Human 5' EST secreted protein SEQ ID NO:457.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
```

```
forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IBL222.
PR 01-AUG-1997; US-905135.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR N-PSDB: X41259.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 27; Page 763-764; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC chemokine activity, haemostatic and thrombolytic activity, chemotactic/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 148 AA.

Query Match 46.4%; Score 667.5; DB 1; Length 148;
Best Local Similarity 83.3%; Pred. No. 9e-60;
Matches 125; Conservative 4; Mismatches 18; Indels 3; Gaps 2;

QY 63 MAAAWSPGSAPAVTARLVGLVFWFVSVTTGPGAVATSAGGESLKCEDLKVGQYICKD 122
Db 1 MAAAWSPGSAPAVTARLVGLVFWFVSVTTGPGAVATSAGGESLKCEDLKVGQYICKD 60
QY 123 PKINDATQEPVNCNTNTAHVSCFPAPNITCKDSGNETHTFTGNEVGFPPISCRNVNGYS 182
Db 61 PKINDATQEPVNCNTNTAHVSCFPAPNITCKDSGNETHTFTGNEVGFPPISCRNVNGYS 120
QY 183 YKVAVALSLFLWLGADRFVLGY-PALGLL 211
Db 121 YXXQXXVS--FSMWVGSRSILPWPCFV 148

RESULT 4
W94360
ID W94360 standard; Protein: 764 AA.
AC W94360;
DE 13-APR-1999 (first entry)
KW Human Factor B analogue #7.
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;
KW stroke; transplant rejection suppression; myocardial infarction.
OS Homo sapiens.
PN US5869615-A.
PD 09-FEB-1999.
PF 03-JAN-1994; 177109.
PR 03-JAN-1994; US-177109.
PA (UNIW ) UNIV WASHINGTON.
PI Hourcade DE, Oglesby TJ;
DR WPI: 99-152877/13.
PT Factor B analogues - with modified complement-mediated activity
```

PS Claim 18; Column -: 53pp; English.  
CC The present sequence represents a human Factor B analogue. The present  
CC invention describes human factor B analogues (see W94353 to W94390) that  
CC exhibit modified complement-mediated activity in vitro and consist of  
CC the sequence given in W94352, of 764 amino acids, with one or more amino  
CC acids in a short consensus repeat (SCR) domain or a von Willebrand  
CC factor (vWF) domain replaced by one or more amino acids from a  
CC corresponding region of an SCR or vWF domain of a second protein, or  
CC in which a protease domain of native factor B is replaced by a protease  
CC domain of C2. Complement protein analogues having altered activity can  
CC be used to control activation of the complement system and are so are  
CC useful for treating disorders associated with abnormal and/or excessive  
CC complement activation e.g. for treating autoimmune diseases, suppressing  
CC transplant rejection or reducing tissue damage associated with myocardial  
CC infarction and stroke.  
CC N.B. The present sequence is not given in the specification but is  
CC derived from the sequence in W94352 as specified in the claims.  
SQ Sequence 764 AA;

Query Match 6.1%; Score 88; DB 1; Length 764;  
Best Local Similarity 24.6%; Pred. No. 1.9;  
Matches 50; Conservative 22; Mismatches 69; Indels 62; Gaps 12;  
QY 24 GTGLYPMRGPFKNLALLPFLSLPLGG-----GSGGKVS----- 61  
DB 2 GSNLSP-----QLCLMPFILGLSGVTPWLSLAQPGQSCSLEGEIKGGSFRLLOEG 55  
QY 61 SKMAAAMPSSG--PSAPEAVTARLVGLVFSVTTGPMGAVAT-----AGGEESLKC-- 111  
DB 56 SLTTSCPSGFYPVQVTRCTCR-----STGWSLTKTODQKTVKACRAIHCPR 105  
QY 111 -EDLVGVQICKDKINDATQEPVNC--TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
DB 106 PHDFENGWEYWPSPYNNVDSIEIFHCYDGYTLRGA-----NRTCQVNGRWSGQTACDNG 161  
QY 167 VGFFK----PISCRNNGYSYKV 185  
DB 162 AGYCSNPGPIGTRKV-GSOYRL 183

RESULT 5  
W94365  
ID W94365 standard; Protein: 764 AA.  
AC W94365;  
DT 13-APR-1999 (first entry)  
DE Human Factor B analogue #12.  
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;  
KW stroke; transplant rejection suppression; myocardial infarction.  
OS Homo sapiens.  
PS Synthetic.  
PN US5869615-A.  
PD 09-FEB-1999.  
PF 03-JAN-1994; US-177109.  
PR 03-JAN-1994; US-177109.  
PA (UNIW ) UNIV WASHINGTON.  
PI Hourcade DE, Oglesby TJ;  
DR WPI: 99-152877/13.  
PT Factor B analogues - with modified complement-mediated activity  
PS Claim 18; Column -: 53pp; English.  
CC The present sequence represents a human Factor B analogue. The present  
CC invention describes human factor B analogues (see W94353 to W94390) that  
CC exhibit modified complement-mediated activity in vitro and consist of  
CC the sequence given in W94352, of 764 amino acids, with one or more amino  
CC acids in a short consensus repeat (SCR) domain or a von Willebrand  
CC factor (vWF) domain replaced by one or more amino acids from a  
CC corresponding region of an SCR or vWF domain of a second protein, or  
CC in which a protease domain of native factor B is replaced by a protease  
CC domain of C2. Complement protein analogues having altered activity can  
CC be used to control activation of the complement system and are so are  
CC useful for treating disorders associated with abnormal and/or excessive  
CC complement activation e.g. for treating autoimmune diseases, suppressing  
CC transplant rejection or reducing tissue damage associated with myocardial

CC infarction and stroke.  
CC N.B. The present sequence is not given in the specification but is  
CC derived from the sequence in W94352 as specified in the claims.  
SQ Sequence 764 AA;

Query Match 6.0%; Score 86; DB 1; Length 764;  
Best Local Similarity 24.6%; Pred. No. 3;  
Matches 50; Conservative 21; Mismatches 70; Indels 62; Gaps 12;  
QY 24 GTGLYPMRGPFKNLALLPFLSLPLGG-----GSGGKVS----- 61  
DB 2 GSNLSP-----QLCLMPFILGLSGVTPWLSLAQPGQSCSLEGEIKGGSFRLLOEG 55  
QY 61 SKMAAAMPSSG--PSAPEAVTARLVGLVFSVTTGPMGAVAT-----SAGGEESLKC-- 111  
DB 56 QALEYVCPSGFYPPVQVTRCTCR-----STGWSLTKTODORSLSKAVCAIHCPR 105  
QY 111 -EDLVGVQICKDKINDATQEPVNC--TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
DB 106 PHDFENGWEYWPSPYNNVDSIEIFHCYDGYTLRGA-----NRTCQVNGRWSGQTACDNG 161  
QY 167 VGFFK----PISCRNNGYSYKV 185  
DB 162 AGYCSNPGPIGTRKV-GSOYRL 183

RESULT 6  
W94372  
ID W94372 standard; Protein: 764 AA.  
AC W94372;  
DT 13-APR-1999 (first entry)  
DE Human Factor B analogue #19.  
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;  
KW stroke; transplant rejection suppression; myocardial infarction.  
OS Homo sapiens.  
PS Synthetic.  
PN US5869615-A.  
PD 09-FEB-1999.  
PF 03-JAN-1994; US-177109.  
PR 03-JAN-1994; US-177109.  
PA (UNIW ) UNIV WASHINGTON.  
PI Hourcade DE, Oglesby TJ;  
DR WPI: 99-152877/13.  
PT Factor B analogues - with modified complement-mediated activity  
PS Claim 18; Column -: 53pp; English.  
CC The present sequence represents a human Factor B analogue. The present  
CC invention describes human factor B analogues (see W94353 to W94390) that  
CC exhibit modified complement-mediated activity in vitro and consist of  
CC the sequence given in W94352, of 764 amino acids, with one or more amino  
CC acids in a short consensus repeat (SCR) domain or a von Willebrand  
CC factor (vWF) domain replaced by one or more amino acids from a  
CC corresponding region of an SCR or vWF domain of a second protein, or  
CC in which a protease domain of native factor B is replaced by a protease  
CC domain of C2. Complement protein analogues having altered activity can  
CC be used to control activation of the complement system and are so are  
CC useful for treating disorders associated with abnormal and/or excessive  
CC complement activation e.g. for treating autoimmune diseases, suppressing  
CC transplant rejection or reducing tissue damage associated with myocardial  
CC infarction and stroke.  
CC N.B. The present sequence is not given in the specification but is  
CC derived from the sequence in W94352 as specified in the claims.  
SQ Sequence 764 AA;

Query Match 6.0%; Score 86; DB 1; Length 764;  
Best Local Similarity 24.1%; Pred. No. 3;  
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;  
QY 24 GTGLYPMRGPFKNLALLPFLSLPLGG-----GSGGKVS----- 61  
DB 2 GSNLSP-----QLCLMPFILGLSGVTPWLSLAQPGQSCSLEGEIKGGSFRLLOEG 55

QY 61 SKMAAMPSPG--PSAPEAVTARLVGLWFVSTTGPACAVATS-----AGGEESLKC-- 111  
Db 56 QALEYVCPGFGYPVPVQTRCR-----STGWSLTKTQDQKTVRKAECHRAHCPR 105  
QY 111 -EDLVQGYCKDKPKINDATQEPVNC-TNTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
Db 106 PHDFENGEXWPRSPYNNVDSIEFHCYDGYTLRCSA---NRTCPRNCMSGQTAICDNG 161  
QY 167 VGFFK-----PISCRNVNGSYKVK 185  
Db 162 AGYCSNPGIPIGTRKV-GSQYRL 183

RESULT 7  
W05153  
ID W05153 standard; Protein; 338 AA.  
AC W05153;  
DT 22-MAY-1997 (first entry)  
DE Rat LAMP residues 1-332.  
KW limbic system associated membrane protein; LAMP; self binding domain;  
KW antibody-like; cell surface adhesion protein; neuron;  
KW monoclonal antibody; 2G9; growth; differentiation; epilepsy;  
KW Alzheimer's disease; schizophrenia; neural stem cell.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT peptide 1..28  
FT protein /note= "Signal peptide"  
FT protein /note= "Mature protein"  
PN W09630052-A1.  
PD 03-OCT-1996.  
PF 29-MAR-1996; U04397.  
PR 31-MAR-1995; US-414657.  
PA (UMDJ-) UMDNJ UNIV NEW JERSEYS HEALTH SCI.  
PI Fischer I, Levitt PR, Pimenta A, Zhukareva V;  
DR WPI: 96-455009/45.  
DR N-PSDB: T42080.  
PT DNA encoding limbic system associated membrane protein self binding  
PT domain - useful in treatment of excessive neural growth in limbic  
PT system, e.g. in animal having epilepsy, Alzheimer's disease or  
PT schizophrenia  
PS Claim 7; Page 40-42; 87pp; English.  
CC The sequences given in W05152-70 represent fragments of limbic system  
CC associated membrane protein (LAMP) self binding domain. LAMP is a  
CC self-binding, antibody-like cell surface adhesion protein, which causes  
CC the formation of connections between adjacent neurons. LAMP is bound by  
CC the monoclonal antibody 2G9 and is thought to be involved in the growth  
CC and differentiation of certain neurons. The protein is highly conserved  
CC and the human and rat sequences differ in only four amino acids. LAMP  
CC nucleic acids, or soluble LAMP analogues, can be used to treat an  
CC animal with excessive neural growth in the limbic region, i.e. where  
CC the animal has epilepsy, Alzheimer's disease or schizophrenia. Neural  
CC stem cells transformed with an expression vector comprising one of  
CC these nucleic acids, can be used to treat neuropathologies involving  
CC the limbic system.  
SQ Sequence 338 AA;

Query Match 5.9%; Score 85; DB 1; Length 338;  
Best Local Similarity 27.7%; Pred. No. 1.2;  
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;  
QY 101 SAGGEESLKCEDLVKVG---QYICKDKPKINDATQEPVNCNTYTAHVSCFPAPNITCKDSS 156  
Db 230 TTGRQASLKCEASVAPDFEWYRDDTRINSANGLEIKSTE---GQSSLVTWNTVEHY 285  
QY 157 GNETHFTGNEVG-----FFKPISCRNVNGSYKVAVALSLFLGWLGADRFLYGLYPALG 209  
Db 286 GNYTCVAANKLGVTNLSLVLPFGSVRGING-SISLAVPL-----WL-----LA 328  
QY 210 LLKFCITVGFC 219  
Db 329 ASLFCLLSKC 338

RESULT 8  
W05154  
ID W05154 standard; Protein; 338 AA.  
AC W05154;  
DT 22-MAY-1997 (first entry)  
DE Rat LAMP residues 1-332.  
KW limbic system associated membrane protein; LAMP; self binding domain;  
KW antibody-like; cell surface adhesion protein; neuron;  
KW monoclonal antibody; 2G9; growth; differentiation; epilepsy;  
KW Alzheimer's disease; schizophrenia; neural stem cell.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT peptide 1..28  
FT protein /note= "Signal peptide"  
FT protein /note= "Mature protein"  
PN W09630052-A1.  
PD 03-OCT-1996.  
PF 29-MAR-1996; U04397.  
PR 31-MAR-1995; US-414657.  
PA (UMDJ-) UMDNJ UNIV NEW JERSEYS HEALTH SCI.  
PI Fischer I, Levitt PR, Pimenta A, Zhukareva V;  
DR WPI: 96-455009/45.  
DR N-PSDB: T42080.  
PT DNA encoding limbic system associated membrane protein self binding  
PT domain - useful in treatment of excessive neural growth in limbic  
PT system, e.g. in animal having epilepsy, Alzheimer's disease or  
PT schizophrenia  
PS Claim 4; Page 43-45; 87pp; English.  
CC The sequences given in W05152-70 represent fragments of limbic system  
CC associated membrane protein (LAMP) self binding domain. LAMP is a  
CC self-binding, antibody-like cell surface adhesion protein, which causes  
CC the formation of connections between adjacent neurons. LAMP is bound by  
CC the monoclonal antibody 2G9 and is thought to be involved in the growth  
CC and differentiation of certain neurons. The protein is highly conserved  
CC and the human and rat sequences differ in only four amino acids. LAMP  
CC nucleic acids, or soluble LAMP analogues, can be used to treat an  
CC animal with excessive neural growth in the limbic region, i.e. where  
CC the animal has epilepsy, Alzheimer's disease or schizophrenia. Neural  
CC stem cells transformed with an expression vector comprising one of  
CC these nucleic acids, can be used to treat neuropathologies involving  
CC the limbic system.  
SQ Sequence 338 AA;

Query Match 5.9%; Score 85; DB 1; Length 338;  
Best Local Similarity 27.7%; Pred. No. 1.2;  
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;  
QY 101 SAGGEESLKCEDLVKVG---QYICKDKPKINDATQEPVNCNTYTAHVSCFPAPNITCKDSS 156  
Db 230 TTGRQASLKCEASVAPDFEWYRDDTRINSANGLEIKSTE---GQSSLVTWNTVEHY 285  
QY 157 GNETHFTGNEVG-----FFKPISCRNVNGSYKVAVALSLFLGWLGADRFLYGLYPALG 209  
Db 286 GNYTCVAANKLGVTNLSLVLPFGSVRGING-SISLAVPL-----WL-----LA 328  
QY 210 LLKFCITVGFC 219  
Db 329 ASLFCLLSKC 338

RESULT 9  
W05156  
ID W05156 standard; Protein; 310 AA.  
AC W05156;  
DT 22-MAY-1997 (first entry)  
DE Rat mature LAMP.  
KW limbic system associated membrane protein; LAMP; self binding domain;  
KW antibody-like; cell surface adhesion protein; neuron;  
KW monoclonal antibody; 2G9; growth; differentiation; epilepsy;





Best Local Similarity 24.1%; Pred. No. 4.7;  
Matches 49; Conservative 22; Mismatches 70; Indels 62; Gaps 12;

QY 24 GTGLYPMRGPFKNLALPFLSLGSG-----GSGGEKVS----- 61  
DB 2 GSNLSP-----QLCLMPFLLGSGVTPWNSLAQPGSCSLEGEIKGSRLLQEG 55

QY 61 SKMAAAPSG--PSAPEAVTARLVGLVFWFVSITGPGAVATS-----AGGESLKC-- 111  
DB 56 QALEYVCPGSGFYPPVQTRTCR-----STGWSLTKTDQKTVRAECPVRCPR 105

QY 111 -EDLKVGQYICKDPKINDATQEPVNC--TNTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
DB 106 PHDFENGGEYWPSPYVNSDEIFCHYDGYTLRGS-----NRTCQVNGRWSGQTATCDNG 161

QY 167 VGFEK-----PISCRNVNGYSYKV 185  
DB 162 AGYCSNPGIPIGTRKV--GSQYRL 183

RESULT 12  
W94373  
ID W94373 standard; Protein; 764 AA.  
AC W94373;  
DT 13-APR-1999 (first entry)  
DE Human Factor B analogue #20.  
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;  
stroke; transplant rejection suppression; myocardial infarction.  
OS Homo sapiens.  
PN US5869615-A.  
PD 03-FEB-1999.  
PF 03-JAN-1994; 177109.  
PR 03-JAN-1994; US-177109.  
PA (UNIW ) UNIV WASHINGTON.  
PI Hourcade DE, Oglesby TJ;  
DR WPI; 99-152877/13.  
PT Factor B analogues - with modified complement-mediated activity  
PS Claim 18; Column -; 53pp; English.  
CC The present sequence represents a human Factor B analogue. The present  
invention describes human Factor B analogues (see W94353 to W94390) that  
exhibit modified complement-mediated activity in vitro and consist of  
the sequence given in W94352, of 764 amino acids, with one or more amino  
acids in a short consensus repeat (SCR) domain or a von Willebrand  
Factor (vWF) domain replaced by one or more amino acids from a  
corresponding region of an SCR or vWF domain of a second protein, or  
in which a protease domain of native Factor B is replaced by a protease  
domain of C2. Complement protein analogues having altered activity can  
be used to control activation of the complement system and are so are  
useful for treating disorders associated with abnormal and/or excessive  
complement activation e.g. for treating autoimmune diseases, suppressing  
transplant rejection or reducing tissue damage associated with myocardial  
infarction and stroke.  
CC N.B. The present sequence is not given in the specification but is  
derived from the sequence in W94352 as specified in the claims.  
SQ Sequence 764 AA;

Query Match 5.8%; Score 84; DB 1; Length 764;  
Best Local Similarity 24.6%; Pred. No. 4.7;  
Matches 50; Conservative 20; Mismatches 71; Indels 62; Gaps 12;

QY 24 GTGLYPMRGPFKNLALPFLSLGSG-----GSGGEKVS----- 61  
DB 2 GSNLSP-----QLCLMPFLLGSGVTPWNSLAQPGSCSLEGEIKGSRLLQEG 55

QY 61 SKMAAAPSG--PSAPEAVTARLVGLVFWFVSITGPGAVATS-----AGGESLKC-- 111  
DB 56 QALEYVCPGSGFYPPVQTRTCR-----STGWSLTKTDQKTVRAECPVRCPR 105

QY 111 -EDLKVGQYICKDPKINDATQEPVNC--TNTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
DB 106 PHDFENGGEYWPSPYVNSDEIFCHYDGYTLRGS-----NRTCQVNGRWSGQTATCDNG 161

QY 167 VGFEK-----PISCRNVNGYSYKV 185  
DB 162 AGYCSNPGIPIGTRKV--GSQYRL 183

RESULT 13  
W94362  
ID W94362 standard; Protein; 763 AA.  
AC W94362;  
DT 13-APR-1999 (first entry)  
DE Human Factor B analogue #9.  
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;  
stroke; transplant rejection suppression; myocardial infarction.  
OS Homo sapiens.  
PN US5869615-A.  
PD 09-FEB-1999.  
PF 03-JAN-1994; 177109.  
PR 03-JAN-1994; US-177109.  
PA (UNIW ) UNIV WASHINGTON.  
PI Hourcade DE, Oglesby TJ;  
DR WPI; 99-152877/13.  
PT Factor B analogues - with modified complement-mediated activity  
PS Claim 18; Column -; 53pp; English.  
CC The present sequence represents a human Factor B analogue. The present  
invention describes human Factor B analogues (see W94353 to W94390) that  
exhibit modified complement-mediated activity in vitro and consist of  
the sequence given in W94352, of 764 amino acids, with one or more amino  
acids in a short consensus repeat (SCR) domain or a von Willebrand  
Factor (vWF) domain replaced by one or more amino acids from a  
corresponding region of an SCR or vWF domain of a second protein, or  
in which a protease domain of native Factor B is replaced by a protease  
domain of C2. Complement protein analogues having altered activity can  
be used to control activation of the complement system and are so are  
useful for treating disorders associated with abnormal and/or excessive  
complement activation e.g. for treating autoimmune diseases, suppressing  
transplant rejection or reducing tissue damage associated with myocardial  
infarction and stroke.  
CC N.B. The present sequence is not given in the specification but is  
derived from the sequence in W94352 as specified in the claims.  
SQ Sequence 763 AA;

Query Match 5.8%; Score 83.5; DB 1; Length 763;  
Best Local Similarity 24.6%; Pred. No. 5.3;  
Matches 49; Conservative 24; Mismatches 71; Indels 55; Gaps 11;

QY 24 GTGLYPMRGPFKNLALPFLSLGSGSGGEKVSVMKMAAAPSGPSAPEAV-----T 78  
DB 2 GSNLSP-----QLCLMPFLLGSLGG-----VTTTPWSLAQPGSCSLEGEIKGGS 48

QY 79 ARLV--GVLFVFSVT---TGPGAVATSAGGESLKCDELKV----- 116  
DB 49 FRLLQEGALEYVCPGSGFYPPASRLCRSTGWSLTKTDQKTVRAECPVRCPRHDF 108

QY 116 --GOYICKDPKINDATQEPVNC--TNTAHVSCFPAPNITCKDSS--GNETHFTGNEGVFF 170  
DB 109 ENGEYWPSPYVNSDEIFCHYDGYTLRGS-----NRTCQVNGRWSGQTATCDNGAGYC 164

QY 171 K-----PISCRNVNGYSYKV 185  
DB 165 SNPGIPIGTRKV--GSQYRL 182

RESULT 14  
W94383  
ID W94383 standard; Protein; 764 AA.  
AC W94383;  
DT 13-APR-1999 (first entry)  
DE Human Factor B analogue #30.  
KW Human; Factor B; variant; mutant; autoimmune disease; tissue damage;  
stroke; transplant rejection suppression; myocardial infarction.



---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2000, 05:16:36 ; Search time 32.32 Seconds  
(without alignments)  
248.566 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 269  
Sequence: 1 MHILKSPNVPRAHGOKNT.....TRLRLSITNETFRKTQLYP 269

Scoring table: OLIGO  
Searched: 82229 seqs, 29864866 residues

base : SwissProt\_38:\*

Word size : 0

Number of hits that pass the threshold : 82229

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.0	1520	1 ABL_DROME	P05222 drosophila
2	8	3.0	1174	1 CRKE_DROME	Q02280 drosophila
3	8	3.0	118	1 CRYP_CRYPA	P52753 cryptoneutr
4	8	3.0	473	1 HN3A_HUMAN	P55317 homo sapien
5	8	3.0	358	1 KRIA_HUMAN	Q05032 homo sapien
6	8	3.0	613	1 XECT_HUMAN	P36021 homo sapien
7	8	3.0	427	1 YEAH_ECOLI	P76235 escherichia
8	8	3.0	753	1 YJ05_CAEEL	Q01975 caenorhabdi
9	7	2.6	1039	1 AG43_ECOLI	P39180 escherichia
10	7	2.6	1959	1 AGRI_RAT	P25304 rattus norv
11	7	2.6	882	1 AREA_ASPNG	O13412 aspergillus
12	7	2.6	550	1 CCF_DROME	P41046 drosophila
13	7	2.6	581	1 CE42_ECOLI	P04419 escherichia
14	7	2.6	551	1 CE46_ECOLI	P00646 escherichia
15	7	2.6	551	1 CE46_ECOLI	P17999 escherichia
16	7	2.6	582	1 CE49_ECOLI	P09883 escherichia
17	7	2.6	510	1 CE4B_ECOLI	P05819 escherichia
18	7	2.6	663	1 CNG2_BOVIN	Q03041 bos taurus
19	7	2.6	424	1 COAA_BPFD	P03661 bacterioph
20	7	2.6	424	1 COAA_BPFD	P03662 bacterioph
21	7	2.6	419	1 CYB_RHOVI	P81378 rhodopsudo
22	7	2.6	2038	1 FSH_DROME	P13709 drosophila
23	7	2.6	907	1 GLR1_MOUSE	P23818 mus musculu
24	7	2.6	572	1 GPC5_HUMAN	P78333 homo sapien
25	7	2.6	183	1 GRP2_ORYSA	P29834 oryza sativ
26	7	2.6	173	1 IPRY_HELPY	P56153 helicobacte
27	7	2.6	575	1 IPR1_SCHPO	Q10286 schizosacch
28	7	2.6	622	1 KICI_HUMAN	P35527 homo sapien
29	7	2.6	166	1 K2C5_BOVIN	P04262 bos taurus
30	7	2.6	796	1 KE3C_MOUSE	O35066 mus musculu
31	7	2.6	638	1 LIK2_RAT	P53670 rattus norv
32	7	2.6	1192	1 LMG2_MOUSE	Q61092 mus musculu
33	7	2.6	534	1 ML51_HUMAN	O15234 homo sapien
34	7	2.6	399	1 MPK4_HUMAN	P45985 homo sapien
35	7	2.6	577	1 MPK4_NEUCR	P23955 neurospora
36	7	2.6	482	1 MPFA_YEAST	P11914 saccharomyc
37	7	2.6	375	1 PER_DROSC	P91607 drosophila
38	7	2.6	375	1 PER_DROSC	P91705 drosophila
39	7	2.6	435	1 PGLX_ASPTU	Q00293 aspergillus

40	7	2.6	623	1 PNT1_DROME	P51022 drosophila
41	7	2.6	718	1 PNT2_DROME	P51023 drosophila
42	7	2.6	3080	1 POLG_ZYMCV	P18479 z genome po
43	7	2.6	163	1 PTFB_BACSU	P26380 bacillus su
44	7	2.6	307	1 RB33_CAEEL	Q20365 caenorhabdi
45	7	2.6	979	1 RFX1_HUMAN	P22670 homo sapien

ALIGNMENTS

RESULT 1

ID	ABL_DROME	STANDARD	PRT	1520 AA
AC	P00522			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	TYROSINE-PROTEIN KINASE DASH/ABL (EC 2.7.1.112).			
GN	DASH OR ABL-1.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 88174728.			
RA	HENKEMEYER M.J., BENNETT R.L., GERTLER F.B., HOFFMANN F.M.;			
RT	"DNA sequence, structure, and tyrosine kinase activity of the			
RT	Drosophila melanogaster Abelson proto-oncogene homolog.,";			
RL	Mol. Cell. Biol. 8:843-853(1988).			
RN	[2]			
RP	SEQUENCE OF 374-648 FROM N.A.			
RX	MEDLINE: 84082064.			
RA	HOFFMANN F.M., FRESCO L.D., HOFFMAN-FALK H., SHILO B.-Z.;			
RT	"Nucleotide sequences of the Drosophila src and abl homologs:			
RT	conservation and variability in the src family oncogenes.,";			
RL	Cell 35:393-401(1983).			
CC	-!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +			
CC	PROTEIN TYROSINE PHOSPHATE.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC			
CC	DOMAIN. BELONGS TO THE ABL SUBFAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M19692; AAA28934.1;			
DR	EMBL: M19690; AAA28934.1; JOINED.			
DR	EMBL: M19691; AAA28934.1; JOINED.			
DR	PIR: A28128; TVFFA.			
DR	HSSP: P00519; 1AB2.			
DR	FLYBASE: FBgn0000017; ABL.			
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.			
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE: PS50001; SH2; 1.			
DR	PROSITE: PS50002; SH3; 1.			
DR	PFAM: PF00017; SH2; 1.			
DR	PFAM: PF00018; SH3; 1.			
DR	PFAM: PF00019; SH3; 1.			
DR	PFAM: PF00069; pkinase; 1.			
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;			
KW	SH2 domain; SH3; 265			
FT	DOMAIN 204 265			
FT	DOMAIN 271 363			

FT DOMAIN 388 644 PROTEIN KINASE.  
FT NP\_BIND 394 402 ATP (BY SIMILARITY).  
FT BINDING 417 417 ATP (BY SIMILARITY).  
FT ACT\_SITE 509 509 BY SIMILARITY.  
FT MOD\_RES 539 539 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CONFLICT 374 377 LSPF -> ASAO (IN REF. 2).  
FT CONFLICT 645 648 ESS1 -> VGDV (IN REF. 2).  
SQ SEQUENCE 1520 AA; 161836 MW; AB831F46 CRC32;

Query Match 3.0%; Score 8; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
| | | | | | | |  
Db 78 GGGGSGSG 85

LT 2  
CIRP\_DROME STANDARD; PRT; 1174 AA.  
ID C1KE\_DROME  
AC Q02280;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POTASSIUM CHANNEL PROTEIN EAG.  
GN EAG.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91262635.

RA WARMKE J., DRYSDALE R., GANETZKY B.;  
RT "A distinct potassium channel polypeptide encoded by the Drosophila  
RT eag locus.";  
RL Science 252:1560-1562(1991).

CC -!- FUNCTION: PROTEIN EAG IS MOST PROBABLY A STRUCTURAL COMPONENT OF  
CC THE POTASSIUM CHANNEL AND MEDIATES THE POTASSIUM PERMEABILITY OF  
CC MEMBRANES.

CC -!- SUBUNIT: DIMER (POTENTIAL).

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND  
CC IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
CC EVERY THIRD POSITION.

CC -!- MISCELLANEOUS: THE SEGMENT H5 IS THOUGHT TO LINE THE CHANNEL PORE.

CC -!- SIMILARITY: TO THE MEMBERS OF THE POTASSIUM CHANNEL PROTEINS  
CC OF THE SH SUPERFAMILY.

CC -!- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; M61157; AAA28495.1;  
CC DR PIR; A40853; A40853.

CC FLYBASE; FBgn0000535; eag.  
CC DR PFAM; PF00027; CNP\_binding; 1.  
CC DR PFAM; PF00914; CNP\_membrane; 1.

CC KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
CC Glycoprotein; Phosphorylation; Repeat.

CC REPEAT 18 91 PAS-1.  
CC FT DOMAIN 116 156 PAC MOTIF.

CC FT TRANSMEM 227 246 SEGMENT S1 (POTENTIAL).  
CC FT TRANSMEM 269 291 SEGMENT S2 (POTENTIAL).  
CC FT TRANSMEM 314 335 SEGMENT S3 (POTENTIAL).  
CC FT TRANSMEM 343 369 SEGMENT S4 (POTENTIAL).

FT TRANSMEM 372 393 SEGMENT S5 (POTENTIAL).  
FT TRANSMEM 442 467 SEGMENT H5 (POTENTIAL).  
FT TRANSMEM 471 493 SEGMENT S6 (POTENTIAL).  
FT CARBOHYD 262 262 POTENTIAL.  
FT CARBOHYD 412 412 POTENTIAL.  
FT CARBOHYD 424 424 POTENTIAL.  
SQ SEQUENCE 1174 AA; 126236 MW; EF855F66 CRC32;

Query Match 3.0%; Score 8; DB 1; Length 1174;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
| | | | | | | |  
Db 1125 GGGGSGSG 1132

RESULT 3

CIRP\_CRYPA STANDARD; PRT; 118 AA.

ID CRYP\_CRYPA

AC P52753;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1999 (Rel. 39, Last annotation update)

DE CRYPARIN PRECURSOR.

GN CRP.

OS Cryphonectria parasitica (Chestnut blight fungus) (Endothia

OS parasitica)

OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;

OC Diaporthales; Valsaceae; Cryphonectria.

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-73.

RX STRAIN-155/2;

RA MEDLINE; 94156182.

RA ZHANG L., VILLALON D., SUN Y., KAZMIERCZAK P., VAN ALFEN N.K.;

RT "Virus-associated down-regulation of the gene encoding cryparin, an

RT abundant cell-surface protein from the chestnut blight fungus,

RT Cryphonectria parasitica.";

RL Gene 139:59-64(1994).

CC -!- FUNCTION: CONTRIBUTES TO SURFACE HYDROPHOBICITY, WHICH IS

CC IMPORTANT FOR PROCESSES SUCH AS ASSOCIATION OF HYPAE IN

CC REPRODUCTIVE STRUCTURES, DISPERSAL OF AERIAL SPORES AND ADHESION

CC OF PATHOGENS TO HOST STRUCTURES. PRODUCED ABUNDANTLY, EXCEPT IN

CC THE DS-RNA VIRUS-INFECTED STRAINS, WHERE THE EXPRESSION IS MUCH

CC REDUCED

CC -!- SUBCELLULAR LOCATION: CELL WALL OF AERIAL HYPAE AND SPOREATION

CC STRUCTURES. ABUNDANTLY SECRETED.

CC -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED ON DAY 2 AND 3 AFTER

CC INOCULATION. A TIME WHEN THE FUNGUS IS IN A RAPID PHASE OF GROWTH.

CC AFTER A STATIONARY PHASE ON DAY 4, THE EXPRESSION DECREASES.

CC -!- SIMILARITY: BELONGS TO THE CERATO-ULMIN HYDROPHOBIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; L09559; AAA19638.1;  
CC DR Cell wall; Signal; Repeat.

CC SIGNAL 1 22

CC CHAIN 23 118 CRYPARIN.

CC DOMAIN 23 32 POLY-GLY.

CC REPEAT 29 42 7 X 2 AA TANDEM REPEAT OF S-G.

CC REPEAT 29 30 1.

CC REPEAT 31 32 2.

CC REPEAT 33 34 3.

CC REPEAT 35 36 4.

CC REPEAT 37 38 5.

CC REPEAT 39 40 6.

```
FT REPEAT 41 42 7. 5B2C48A6 CRC32;
SQ SEQUENCE 118 AA; 11387 MW; 11387 MW; 5B2C48A6 CRC32;

Query Match 3.0%; Score 8; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGG 55
Db 25 GGGGSGG 32

RESULT 4
HN3A_HUMAN STANDARD; PRT; 473 AA.
ID HN3A_HUMAN
AC P5317;
DT 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DL 01-NOV-1997 (Rel. 35, Last annotation update)
CC HEPATOCYTE NUCLEAR FACTOR 3-ALPHA (HNF-3A).
CC HNF3A OR TCF3A.
CC Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96254057.
RA BINGLE C.D., GOWAN S.;
RT "Molecular cloning of the forkhead transcription factor HNF-3 alpha
RT from a human pulmonary adenocarcinoma cell line.";
RL Biochim. Biophys. Acta 1307:17-20(1996).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES
CC SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PEPCCK, ETC.
CC INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL; U39840; AAB06493.1;
MIN; 602294;
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR PFAM; PF00250; Fork_head; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DNA_BIND 170 261 FORK-HEAD.
SQ SEQUENCE 473 AA; 49321 MW; AB4F7931 CRC32;

Query Match 3.0%; Score 8; DB 1; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGSGG 55
Db 277 GGGGSGG 284

RESULT 5
KKIA_HUMAN STANDARD; PRT; 358 AA.
ID KKIA_HUMAN
AC Q00532;
DT 01-APR-1993 (Rel. 25, Created)
DL 01-APR-1993 (Rel. 25, Last sequence update)
CC -!- SIMILARITY: BELONGS TO THE MONOCARBOXYLATE TRANSPORTER FAMILY.
CC -!- SIMILARITY: BELONGS TO THE MONOCARBOXYLATE TRANSPORTER FAMILY.
DT 01-APR-1993 (Rel. 25, Last sequence update)

15-JUL-1999 (Rel. 38, Last annotation update)
SERINE/THREONINE-PROTEIN KINASE KKIALRE (EC 2.7.1.1).
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92347325.
RA HARLOW E., TSAI L.-H.;
RT "A family of human cdc2-related protein kinases.";
RL EMBO J. 11:2909-2917(1992).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC2/CDKX SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL; X66358; CAA47002.1;
EMBL; X66359; CAA47002.1; JOINED.
PIR; S22744; S22744.
DR PIR; S22745; S22745.
DR PIR; S23383; S23383.
DR HSP; P24941; 1A01.
DR MIM; 603441;
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 5 288 PROTEIN KINASE.
FT NP_BIND 11 19 ATP (BY SIMILARITY).
FT BINDING 34 34 ATP (BY SIMILARITY).
FT ACT_SITE 127 127 BY SIMILARITY.
SQ SEQUENCE 358 AA; 41834 MW; 87B81E2E CRC32;

Query Match 3.0%; Score 8; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 YPALGLLK 212
Db 259 YPALGLLK 266

RESULT 6
XPCT_HUMAN STANDARD; PRT; 613 AA.
ID XPCT_HUMAN
AC P36021;
DT 01-JUN-1994 (Rel. 29, Created)
DL 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE X-LINKED PEST-CONTAINING TRANSPORTER.
GN SLIC16A2 OR XPCT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95072579.
RA LAPRENIERE R.G., CARREL L., WILLARD H.F.;
RT "A novel transmembrane transporter encoded by the XPCT gene in
RT Xq13.2.";
RL Hum. Mol. Genet. 3:1133-1140(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE MONOCARBOXYLATE TRANSPORTER FAMILY.
CC -!- SIMILARITY: BELONGS TO THE MONOCARBOXYLATE TRANSPORTER FAMILY.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U05321; AAB60375.1; -  
CC EMBL; U05316; AAB60375.1; JOINED.  
CC EMBL; U05317; AAB60375.1; JOINED.  
CC EMBL; U05318; AAB60375.1; JOINED.  
CC EMBL; U05319; AAB60375.1; JOINED.  
CC EMBL; U05320; AAB60375.1; JOINED.  
CC EMBL; U05315; AAB60374.1; -  
CC MIM; 300095; -  
CC PFAM; PF01587; MCT; 1.  
KW Transport; Transmembrane; Glycoprotein.  
DOMAIN 1 70 GLY/SER-RICH.  
TRANSMEM 171 191 POTENTIAL.  
TRANSMEM 218 238 POTENTIAL.  
TRANSMEM 246 266 POTENTIAL.  
TRANSMEM 275 295 POTENTIAL.  
TRANSMEM 304 324 POTENTIAL.  
TRANSMEM 333 353 POTENTIAL.  
TRANSMEM 397 417 POTENTIAL.  
TRANSMEM 431 451 POTENTIAL.  
TRANSMEM 461 481 POTENTIAL.  
TRANSMEM 484 504 POTENTIAL.  
TRANSMEM 522 542 POTENTIAL.  
TRANSMEM 552 572 POTENTIAL.  
SQ SEQUENCE 613 AA; 6640 MW; D5191340 CRC32;

Query Match 3.0%; Score 8; DB 1; Length 613;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
| | | | | | | |  
DB 34 GGGGSGSG 41

RESULT 7  
YEAR\_ECOLI  
ID YEAH\_ECOLI STANDARD; PRT; 427 AA.  
P76235; 007964;  
15-JUL-1998 (Rel. 36, Created)  
15-JUL-1998 (Rel. 36, Last sequence update)  
15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 49.4 KD PROTEIN IN GAPA-RND INTERGENIC REGION.  
GN YEAH  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MGL655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE; 97251358.  
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,  
RA KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,  
RA MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,  
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,

RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,  
RA YAMAMOTO Y., HORIUCHI T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 40.1-50.0 min region on the linkage map";  
RL DNA Res. 3:379-392(1996).  
CC -1- SIMILARITY: TO B.SUBTILIS YHBH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE00273; AAC74854.1; -  
CC EMBL; D90822; CAB21509.1; -  
CC EGOGENE; EG13494; YEAH.  
KW Hypothetical protein.  
SQ SEQUENCE 427 AA; 49392 MW; 42F92C9E CRC32;

Query Match 3.0%; Score 8; DB 1; Length 427;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
| | | | | | | |  
DB 94 GGGGSGSG 101

RESULT 8  
YJ05\_CAEEL  
ID YJ05\_CAEEL STANDARD; PRT; 753 AA.  
AC 001975;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL 85.1 KD NUCLEASE C41D11.5 (EC 3.1.30.-).  
GN C41D11.5  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA GARTUNG S., MAGGI L.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE DNA/RNA NON-SPECIFIC ENDONUCLEASES  
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF003740; AAC48141.1; -  
CC WORMPEP; C41D11.5; CE08662.  
DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.  
DR PFAM; PF01223; Endonuclease; 1.  
KW Hypothetical protein; Hydrolase; Nuclease; Endonuclease.  
FT ACT\_SITE 593 593 BY SIMILARITY.  
SQ SEQUENCE 753 AA; 85068 MW; 9DF3803A CRC32;

Query Match 3.0%; Score 8; DB 1; Length 753;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 GADRFYLG 204



```

Db 398 GADREYLG 405
|||||||
RESULT 9
AG43_ECOLI STANDARD: PRT: 1039 AA.
AC P39150; P76360; P75614; P97241; Q46771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ANTIGEN 43 PRECURSOR (AG43) (FLUFFING PROTEIN).
GN FLU
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RP [1]
RP SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RP MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.,
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474 (1997).
RN [2]
RN SEQUENCE FROM N.A.
RN STRAIN-K12;
RN MEDLINE; 97251358.
RX ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RX KASAI H., KIMURA S., KITAGAWA M., KITAGAWA K., MAKINO K., MIKI T.,
RX MIZOBUCHI K., MORI H., MORI T., MOMOMURA K., NAKADE S., NAKAMURA Y.,
RX NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RX SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,
RX YAMAMOTO Y., HORIUCHI T.,
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map."
RL DNA Res. 3:379-392 (1996).
RN [3]
RN SEQUENCE FROM N.A.
RN STRAIN-ML 308-225;
RN HENDERSON I.R., OWEN P.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RN PARTIAL SEQUENCE.
RN STRAIN-ML 308-225;
RN MEDLINE; 89291704.
RT "Purification and N-terminal sequence of the alpha subunit of antigen
RT 43, a unique protein complex associated with the outer membrane of
RT Escherichia coli."
RL J. Bacteriol. 171:3634-3640 (1989).
RN [5]
RN SEQUENCE OF 53-63.
RN STRAIN-K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12."
RL Electrophoresis 18:1259-1313 (1997).
RN [6]
RN GENE NAME.
RX MEDLINE; 97257509.
RA HENDERSON I.R., MEEHAN M., OWEN P.;
RT "Antigen 43, a phase-variable bipartite outer membrane protein,
RT determines colony morphology and autoaggregation in Escherichia coli
RT K-12."
RL FEMS Microbiol. Lett. 149:115-120 (1997).
CC -!- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION.
CC -!- SUBUNIT: CONSIST OF TWO SUBUNITS; ALPHA AND BETA.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -!- SIMILARITY: TO ADHESIN AIDI-I AND TO BORDETTELLA PERTACTIN.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----

EMBL; AE000291; AAC75061.1; ALT\_INIT.  
EMBL; D90838; CAB21751.1; ALT\_INIT.  
EMBL; D90839; CAB21756.1; ALT\_INIT.  
EMBL; U24429; AAB47869.1;  
ECOGENE; EGI2686; FLU.  
KW Outer membrane; Signal.  
FT SIGNAL 1 52  
FT CHAIN 53 551  
FT CHAIN 552 1039  
FT VARIANT 2 2  
FT VARIANT 41 42  
FT VARIANT 46 46  
FT VARIANT 157 157  
FT VARIANT 188 188  
FT VARIANT 303 305  
FT VARIANT 320 320  
FT VARIANT 372 372  
FT VARIANT 493 493  
FT VARIANT 497 497  
FT VARIANT 585 585  
FT VARIANT 709 709  
FT VARIANT 721 721  
FT VARIANT 753 753  
FT VARIANT 801 803  
FT VARIANT 815 815  
FT VARIANT 824 824  
FT VARIANT 829 835  
FT VARIANT 845 847  
FT VARIANT 855 855  
FT VARIANT 888 888  
FT VARIANT 1025 1025  
FT VARIANT 61 63  
FT CONFLICT 824 824  
SQ SEQUENCE 1039 AA; 106841 MW; 49DC5F84 CRC32;

ANTIGEN 43 ALPHA CHAIN.  
ANTIGEN 43 BETA CHAIN.  
K -> N (IN STRAIN ML 308-225).  
SL -> FF (IN STRAIN ML 308-225).  
T -> K (IN STRAIN ML 308-225).  
W -> L (IN STRAIN ML 308-225).  
V -> F (IN STRAIN ML 308-225).  
ATN -> STI (IN STRAIN ML 308-225).  
A -> T (IN STRAIN ML 308-225).  
N -> Q (IN STRAIN ML 308-225).  
E -> V (IN STRAIN ML 308-225).  
S -> N (IN STRAIN ML 308-225).  
H -> Y (IN STRAIN ML 308-225).  
E -> K (IN STRAIN ML 308-225).  
M -> T (IN STRAIN ML 308-225).  
GHL -> SHF (IN STRAIN ML 308-225).  
S -> P (IN STRAIN ML 308-225).  
A -> V (IN STRAIN ML 308-225).  
C -> S (IN STRAIN ML 308-225).  
LNLVHTS -> MNLINA (IN STRAIN ML 308-225).  
OGT -> LGA (IN STRAIN ML 308-225).  
S -> T (IN STRAIN ML 308-225).  
Q -> L (IN STRAIN ML 308-225).  
S -> I (IN STRAIN ML 308-225).  
ETV -> TTT (IN REF. 5).  
C -> S (IN REF. 2).

Query Match 2.6%; Score 7; DB 1; Length 1039;  
Best Local Similarity 100.0%; Pred.No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VAVALS 191  
Db 36 VAVALS 42  
|||||||

RESULT 10  
AGRI\_RAT  
ID AGRI\_RAT STANDARD; PRT: 1959 AA.  
AC P25304; Q63034;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE AGRIN PRECURSOR.  
GN AGRN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.  
RC TISSUE-EMBRYONIC SPINAL CORD;  
RX MEDLINE; 91222570.  
RA RUPP F., PAVAN D.G., MAGILL-SOLC C., COWAN D.M., SCHELLER R.H.;  
RT "Structure and expression of a rat agrin.";

RL Neuron 6:811-823(1991).  
 RN [2]  
 RP SEQUENCE OF 1777-1801 FROM N.A.  
 RX MEDLINE; 92407628.  
 RA RUPP F., OZCELIK T., LINIAL M., PETERSON K., FRANCKE U., SCHELLER R.;  
 RT "Structure and chromosomal localization of the mammalian agrin gene.";  
 RL J. Neurosci. 12:3535-3544(1992).  
 CC -!- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE  
 CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE  
 CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.  
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR  
 CC JUNCTION.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST FIVE DIFFERENT FORMS ARISE BY  
 CC ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR  
 CC CLUSTERING ACTIVITY.  
 CC -!- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.  
 CC -!- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.  
 CC -!- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE  
 CC INITIATOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M64780; AAA40703.1; -;  
 DR EMBL; M64780; AAA40702.1; ALT\_INIT.  
 DR EMBL; S44194; AAB23326.1; -;  
 DR PIR; JH0399; AGRT.  
 DR HSSP; P00740; 11XA.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR PFAM; PF00050; kazal; 9.  
 DR PFAM; PF00053; laminin\_EGF; 2.  
 DR PFAM; PF00054; laminin\_G; 3.  
 DR PFAM; PF01390; SEA; 1.  
 DR PFAM; PF01390; SEA; 1.  
 KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;  
 KW Laminin EGF-like domain.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1959  
 FT DOMAIN 65 137  
 FT KAZAL-LIKE 1.  
 FT DOMAIN 141 212  
 FT KAZAL-LIKE 2.  
 FT DOMAIN 213 284  
 FT KAZAL-LIKE 3.  
 FT DOMAIN 287 356  
 FT KAZAL-LIKE 4.  
 FT DOMAIN 361 429  
 FT KAZAL-LIKE 5.  
 FT DOMAIN 430 494  
 FT KAZAL-LIKE 6.  
 FT DOMAIN 495 559  
 FT KAZAL-LIKE 7.  
 FT DOMAIN 563 645  
 FT KAZAL-LIKE 8.  
 FT DOMAIN 688 741  
 FT LAMININ EGF-LIKE 1.  
 FT DOMAIN 742 788  
 FT LAMININ EGF-LIKE 2.  
 FT DOMAIN 794 864  
 FT KAZAL-LIKE.  
 FT DOMAIN 1220 1258  
 FT EGF-LIKE 1.  
 FT DOMAIN 1440 1477  
 FT EGF-LIKE 2.  
 FT DOMAIN 1479 1516  
 FT EGF-LIKE 3.  
 FT DOMAIN 1709 1748  
 FT EGF-LIKE 4.  
 FT DOMAIN 869 992  
 FT SER/THR-RICH.  
 FT DOMAIN 1147 1215  
 FT SER/THR-RICH.  
 FT POTENTIAL.  
 FT DISULFID 97 116  
 FT DISULFID 105 137  
 FT POTENTIAL.  
 FT DISULFID 171 191  
 FT POTENTIAL.  
 FT DISULFID 180 212  
 FT POTENTIAL.  
 FT DISULFID 244 263  
 FT POTENTIAL.  
 FT DISULFID 252 284  
 FT POTENTIAL.  
 FT DISULFID 316 335  
 FT POTENTIAL.  
 FT DISULFID 324 356  
 FT POTENTIAL.  
 FT DISULFID 389 408  
 FT POTENTIAL.

FT DISULFID 397 429  
 FT DISULFID 454 473  
 FT DISULFID 462 494  
 FT DISULFID 518 538  
 FT DISULFID 527 559  
 FT DISULFID 604 624  
 FT DISULFID 613 645  
 FT DISULFID 688 700  
 FT DISULFID 690 707  
 FT DISULFID 709 718  
 FT DISULFID 721 739  
 FT DISULFID 742 754  
 FT DISULFID 744 761  
 FT DISULFID 763 772  
 FT DISULFID 775 786  
 FT DISULFID 823 843  
 FT DISULFID 832 864  
 FT DISULFID 1224 1235  
 FT DISULFID 1229 1246  
 FT DISULFID 1248 1257  
 FT DISULFID 1444 1455  
 FT DISULFID 1449 1465  
 FT DISULFID 1467 1476  
 FT DISULFID 1483 1494  
 FT DISULFID 1488 1504  
 FT DISULFID 1506 1515  
 FT DISULFID 1713 1727  
 FT DISULFID 1721 1736  
 FT DISULFID 1738 1747  
 FT CARBOHYD 145 145  
 FT CARBOHYD 672 672  
 FT CARBOHYD 827 827  
 FT CARBOHYD 957 957  
 FT VARSPLIC 1144 1152  
 FT VARSPLIC 1780 1798  
 FT VARSPLIC 1788 1798  
 FT VARSPLIC 1780 1787  
 FT VARIANT 314 314  
 SQ SEQUENCE 1959 AA; 208645 MW; 6B6E022E CRC32;  
 Query Match 2.6%; Score 7; DB 1; Length 1959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 50 GSGSGSGE 56  
 DB 567 GSGSGSGE 573  
 RESULT 11  
 AREA ASPNG STANDARD; PRT; 882 AA.  
 ID AREA ASPNG  
 AC O13412;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NITROGEN REGULATORY PROTEIN AREA.  
 GN AREA.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-N402;  
 RA MACCABE A.P., VANHANEN S.A.S., SOLLEWIJN GELPKE M.,  
 RA VAN DE VONDERVOORT P., ARST H.N., VISSER J.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases  
 CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING  
 CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.  
 CC HIGH, TO OTHER FUNGAL NITROGEN REGULATORY PROTEINS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X81998; CAA57524.1; -  
DR HSP; P17429; 5GAT.  
DR PROSITE; PS00344; GATA\_ZN\_FINGER; 1.  
DR PFAM; PF00320; GATA; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
KW Nuclear protein; Nitrate assimilation.  
FT ZN\_FING 676 700 GATA-TYPE.  
SQ SEQUENCE 882 AA; 94518 MW; 54DEA816 CRC32;

Query Match 2.6%; Score 7; DB 1; Length 882;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 LGGGGSG 53  
Db 6 LGGGGSG 12

RESULT 12  
CCF\_DROME STANDARD; PRT; 550 AA.  
AC P41046; Q24271;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CENTROSMAL AND CHROMOSOMAL FACTOR (CCF) (CHROMOCENTROSOMIN).  
GN CORTO.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=OREGON-R;  
RC MEDLINE; 98130597.  
RA KODJABACHIAN L., DELAAGE M., MAUREL C., MIASSOD R., JACQ B.,  
RA ROSSET R.;  
PM "Mutations in ccf, a novel Drosophila gene encoding a chromosomal  
factor, affect progression through mitosis and interact with Pc-G  
mutations.";  
RL EMBO J. 17:1063-1075(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA TAKAMURA C., IMAMURA Y., TAIRA T., IGUCHI-ARIGA S., ARIGA H.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR PROPER CONDENSATION OF MITOTIC CHROMOSOMES  
CC AND PROGRESSION THROUGH MITOSIS. IS AN ESSENTIAL GENE. BINDS TO  
CC SPECIFIC POLYTENE CHROMOSOME SITES, MANY OF WHICH ARE SHARED WITH  
CC THE POSTERIOR SEX COMBS PROTEIN.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CENTROSMAL.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING OOGENESIS, EMBRYONIC AND  
CC LARVAL STAGES.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS  
CC FRAMESHIFTS.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U35074; AAC13917.1; -  
DR EMBL; D43795; BAA07853.1; ALT\_FRAME.  
DR HSP; P04002; IWFA.  
DR FLYBASE; FBgn010313; corto.  
KW DNA-binding; Nuclear protein; Mitosis; Coiled coil.  
FT DOMAIN 26 35 POLY-GLN.  
FT DOMAIN 46 50 POLY-GLN.  
FT DOMAIN 66 72 POLY-GLN.  
FT DOMAIN 84 89 POLY-SER.  
FT DOMAIN 110 126 POLY-GLN.  
FT DOMAIN 204 210 POLY-SER.  
FT DOMAIN 215 220 POLY-GLY.  
FT DOMAIN 236 275 COILED COIL (POTENTIAL).  
FT DOMAIN 243 250 POLY-GLN.  
FT DOMAIN 253 258 POLY-GLN.  
FT DOMAIN 285 297 POLY-GLN.  
FT DOMAIN 322 329 POLY-ALA.  
FT DOMAIN 332 336 POLY-ALA.  
FT DOMAIN 396 410 POLY-GLN.  
FT DOMAIN 414 417 POLY-PRO.  
FT DOMAIN 466 469 POLY-GLY.  
FT DOMAIN 504 507 POLY-ALA.  
FT DOMAIN 512 518 POLY-ALA.  
SQ SEQUENCE 550 AA; 58930 MW; F927C5E0 CRC32;

Query Match 2.6%; Score 7; DB 1; Length 550;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 LGGGGSG 53  
Db 214 LGGGGSG 220

RESULT 13  
CER2\_ECOLI STANDARD; PRT; 581 AA.  
ID CER2\_ECOLI  
AC P04419;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE COLICIN E2 (EC 3.1.21.1).  
GN COL OR CERAB.  
OS Escherichia coli.  
OG Plasmid ColE2-P9.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85239907.  
RA COLE S.T., SAINT-JOANIS B., PUGSLEY A.P.;  
RT "Molecular characterisation of the colicin E2 operon and  
RT identification of its products.";  
RL Mol. Gen. Genet. 198:465-472(1985).  
RN [2]  
RP SEQUENCE OF 377-581 FROM N.A.  
RX MEDLINE; 85062845.  
RA LAU P.C.K., ROWSOME R.W., ZUKER M., VISENTIN L.P.;  
RT "Comparative nucleotide sequences encoding the immunity proteins and  
RT the carboxyl-terminal peptides of colicins E2 and E3.";  
RL Nucleic Acids Res. 12:8733-8745(1984).  
RN [3]  
RP SEQUENCE OF 478-581 FROM N.A.  
RX MEDLINE; 85215569.  
RA MASAKI H., TOBA M., OHTA T.;  
RT "Structure and expression of the Cole2-P9 immunity gene.";  
RL Nucleic Acids Res. 13:1623-1635(1985).  
CC -!- FUNCTION: THIS PLASMID-CODED BACTERICIDAL PROTEIN IS AN  
CC ENDONUCLEASE ACTIVE ON BOTH SINGLE- AND DOUBLE-STRANDED DNA BUT  
CC WITH UNDEFINED SPECIFICITY.  
CC -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE  
CC AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.  
CC -----



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X15856; CAA33855.1; -;  
 DR EMBL; M31808; AAA23080.1; -;  
 KW Plasmid; Bacteriocin; Colicin; Toxin; Hydrolase; Endonuclease.  
 FT DOMAIN 455 551 RIBOSOME INACTIVATING ACTIVITY.  
 FT DOMAIN 530 551 BINDING OF IMMUNITY PROTEIN (BY  
 FT SIMILARITY).  
 SQ SEQUENCE 551 AA; 58011 MW; 34fB6FC9 CRC32;

Query Match 2.6%; Score 7; DB 1; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 49 GGGGGSG 55  
 |||||  
 Db 47 GGGGGSG 53

Search completed: February 11, 2000, 05:25:12  
 Job time: 516 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2000, 05:40:21 ; Search time 41.27 Seconds  
(without alignments)  
451.923 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 269  
Sequence: 1 MHILKSPNVIPRAHGOKNT.....TRLRLSTNTRFKTQLYP 269

Scoring table: OLIGO

Searched: 225878 seqs, 59334122 residues

base : SPTREMBL\_12.\*

Word size : 0

Number of hits that pass the threshold : 225878

- 1: sp.archaea.\*
- 2: sp.bacteria.\*
- 3: sp.fungi.\*
- 4: sp.human.\*
- 5: sp.invertebrate.\*
- 6: sp.mammal.\*
- 7: sp.mhc.\*
- 8: sp.organelle.\*
- 9: sp.phage.\*
- 10: sp.plant.\*
- 11: sp.podent.\*
- 12: sp.virus.\*
- 13: sp.vertebrate.\*
- 14: sp.unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	3.0	262	2	Q44188	Q44188 agrobacteri
2	8	3.0	529	2	P74332	P74332 synechocyst
3	8	3.0	1404	4	O15080	O15080 homo sapien
4	8	3.0	816	5	Q24538	Q24538 drosophila
5	8	3.0	598	5	Q24732	Q24732 drosophila
6	8	3.0	548	5	O19496	O19496 caenorhabdi
7	8	3.0	530	5	O76918	O76918 drosophila
8	8	3.0	255	8	Q9XKD9	Q9XKD9 dicyema mis
9	8	3.0	322	10	O65517	O65517 arabidopsis
10	8	3.0	2946	10	O64634	O64634 arabidopsis
11	8	3.0	207	10	O43522	O43522 lycopersico
12	8	3.0	256	10	O24568	O24568 zea mays (m
13	8	3.0	288	10	O96463	O96463 hordeum vul
14	8	3.0	330	12	O89903	O89903 mouse cytom
15	8	3.0	252	12	O36415	O36415 alcelaphine
16	7	2.6	434	1	O74038	O74038 cenarchaeum
17	7	2.6	317	1	O57894	O57894 pyrococcus
18	7	2.6	411	1	O58003	O58003 pyrococcus
19	7	2.6	679	1	O59078	O59078 pyrococcus
20	7	2.6	510	1	O59101	O59101 pyrococcus
21	7	2.6	852	1	O39DL5	O39DL5 aeropyrum p
22	7	2.6	301	1	O39DA6	O39DA6 aeropyrum p
23	7	2.6	343	1	O9YAA7	O9YAA7 aeropyrum p
24	7	2.6	523	2	O54200	O54200 streptomyce
25	7	2.6	506	2	O31953	O31953 bacillus su

26	7	2.6	293	2	O51095	O51095 borrelia bu
27	7	2.6	458	2	O05591	O05591 mycobacteri
28	7	2.6	273	2	O83115	O83115 treponema p
29	7	2.6	280	2	O83475	O83475 treponema p
30	7	2.6	248	2	O84240	O84240 chlamydia t
31	7	2.6	489	2	O85724	O85724 streptomyce
32	7	2.6	448	2	O24842	O24842 acinetobact
33	7	2.6	242	2	O46297	O46297 clostridium
34	7	2.6	288	2	P73802	P73802 synechocyst
35	7	2.6	696	2	O54201	O54201 streptomyce
36	7	2.6	576	2	O51604	O51604 escherichia
37	7	2.6	173	2	O92LL5	O92LL5 helicobacte
38	7	2.6	308	2	O51635	O51635 shigella so
39	7	2.6	308	2	O51636	O51636 shigella so
40	7	2.6	308	2	O51639	O51639 escherichia
41	7	2.6	466	2	O87185	O87185 streptococc
42	7	2.6	499	2	O9XB47	O9XB47 escherichia
43	7	2.6	168	2	O9X972	O9X972 streptococc
44	7	2.6	640	2	O9WZB2	O9WZB2 thermotoga
45	7	2.6	889	3	O42723	O42723 emericella

#### ALIGNMENTS

```

RESULT 1
Q44188 PRELIMINARY; PRT; 262 AA.
AC Q44188;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.
GN OATA.
OS Agrobacterium radiobacter.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1305 LAC9;
RX MEDLINE; 96423889.
RA TIBURTUS A., DE LUCA N., HUSSAIN H., JOHNSTON A.W.B.;
RT "Expression of the exy gene, required for exopolysaccharide synthesis
  in Agrobacterium, is activated by the regulatory ros gene.";
RL Microbiology 142:2621-2629(1996).
DR EMBL; X95394; CAA64678.1; -.
DR HSSP; P16932; LDGE.
DR PFAM; PF00202; aminotran_3; 1.
KW Transferase.
SQ SEQUENCE 262 AA; 28124 MW; DF7A4568 CRC32;

```

Query Match 3.0%; Score 8; DB 2; Length 262;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 4 LKSPNVI 11
    |||||
Db 244 LKSPNVI 251

```

```

RESULT 2
P74332 PRELIMINARY; PRT; 529 AA.
ID P74332;
AC P74332;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 38.0 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RW EMBL; D90914; BAA18426.1; -.
KW Hypothetical protein.
SQ SEQUENCE 529 AA; 57992 MW; 15123FB3 CRC32;

Query Match 3.0%; Score 8; DB 2; Length 529;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 SLPLGGG 50
Db 260 SLPLGGG 267

RESULT 3
ID O15080 PRELIMINARY; PRT; 1404 AA.
AC O15080;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE KIA0375.
GN KIA0375.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRAIN;
RX MEDLINE; 97349984.
RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RW EMBL; AB002373; BAA20830.1; -.
SQ SEQUENCE 1404 AA; 149264 MW; 675DE431 CRC32;

Query Match 3.0%; Score 8; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 LALLPFSL 44
Db 1163 LALLPFSL 1170

RESULT 4
ID Q24538 PRELIMINARY; PRT; 816 AA.
AC Q24538;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE STEROID RECEPTOR BETA FTZ-F1.

```

```

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91227912.
RA LAVORNA G., UEDA H., CLOS J., WU C.;
RT "FTZ-F1, a steroid hormone receptor-like protein implicated in the
RT activation of fushi tarazu.";
RL Science 252:848-851(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93219413.
RA LAVORNA G., KARIM F.D., THUMMEL C.S., WU C.;
RT "Potential role for a FTZ-F1 steroid receptor superfamily member in
RT the control of Drosophila metamorphosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3004-3008(1993).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; M98397; AAA28915.1; -.
DR HSP; P19793; 2NLL.
DR FLYBASE; FBgn001078; ftz-fl.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PFAM; PF00104; hormone_rec; 1.
DR PFAM; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR KW Hormone; Receptor; Transcription regulation; DNA-binding;
DR KW Nuclear protein; Zinc-finger.
SQ SEQUENCE 816 AA; 87415 MW; 949464D9 CRC32;

Query Match 3.0%; Score 8; DB 5; Length 816;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGGGGSGG 55
Db 96 GGGGGSGG 103

RESULT 5
ID Q24732 PRELIMINARY; PRT; 598 AA.
AC Q24732;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE GLASS.
GN GLASS.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95327686.
RA O'NEILL E.M., ELLIS M.C., RUBIN G.M., TJIAN R.;
RT "Functional domain analysis of glass, a zinc-finger-containing
RT transcription factor in Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6557-6561(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX LIU H., MA C., MOSES K.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39746; AAA85272.1; -.
DR HSP; P08046; IALF.
DR TRANSFAC; T02282; -.
DR FLYBASE; FBgn0015211; Dvir'g1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 5.
DR PFAM; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR KW Zinc-finger; Metal-binding; DNA-binding.
SQ SEQUENCE 598 AA; 61773 MW; 87921C20 CRC32;

```

Query Match 3.0%; Score 8; DB 5; Length 598;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
 DB 244 GGGGSGSG 251

RESULT 6  
 ID Q19496 PRELIMINARY; PRT; 548 AA.  
 AC Q19496;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 63.5 KD PROTEIN.  
 F16H11.5.  
 Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,  
 RA CRAXFORD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WAYSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WU X.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U55376; AAA98005.1; -;  
 DR HSSP; P03372; LHCP.  
 DR PFAM; PF00105; zf-C4; 1.  
 DR PRINTS; PR00047; STROIDFINGER.  
 KW Hypothetical protein.  
 FT DOMAIN 139 144  
 SQ SEQUENCE 548 AA; 63545 MW; D07D3BA5 CRC32;

Query Match 3.0%; Score 8; DB 5; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 KDSSGNET 160  
 DB 109 KDSSGNET 116

RESULT 7  
 ID O76918 PRELIMINARY; PRT; 530 AA.  
 AC O76918;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE EG:96GI0.7 PROTEIN.  
 GN EG:96GI0.7.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BOLSHAKOV V., BORKOVA D., MINANA B., KAFATOS F.;  
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA BENOS P.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL024484; CAA19668.1; -;  
 SQ SEQUENCE 530 AA; 54644 MW; 071B0D78 CRC32;

Query Match 3.0%; Score 8; DB 5; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
 DB 65 GGGGSGSG 72

RESULT 8  
 ID Q9XKD9 PRELIMINARY; PRT; 255 AA.  
 AC Q9XKD9;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CYTOCHROME C OXIDASE SUBUNIT III.  
 GN COIII.  
 OS Dicyema misakiense.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Mesozoa; Rhombozoa; Dicyemida; Dicyemidae;  
 OC Dicyema.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 99150413.  
 RA WATANABE K.I., BESSHO Y., KAWASAKI M., HORI H.;  
 RT "Mitochondrial genes are found on minicircle DNA molecules in the  
 RT mesozoan animal Dicyema.";  
 RL J. Mol. Biol. 286:345-650(1999).  
 DR EMBL; AB011834; BAA76312.1; -;  
 KW Mitochondrion.  
 SQ SEQUENCE 255 AA; 28779 MW; 865DC7A1 CRC32;

Query Match 3.0%; Score 8; DB 8; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 PFSLPLLG 48  
 DB 127 PFSLPLLG 134

RESULT 9  
 ID O65517 PRELIMINARY; PRT; 322 AA.  
 AC O65517;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 34.1 KD PROTEIN.  
 GN F23E13.150.



OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., HILBERT H., BRAUN M., HOLZER E., BRANDT A., DUESTERHOEFT A.,  
 RA HOEISEL J., JESSE T., HELJNEN L., VOS P., MEWES H.W., MAYER K.,  
 RA SCHUELLER C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL022141; CAA18132.1; -  
 DR MENDEL; 29098; Arath; 2109; 29098.  
 KW Hypothetical protein.  
 SQ SEQUENCE 322 AA; 34144 MW; D8212ADB CRC32;  
 Query Match 3.0%; Score 8; DB 10; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 7.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 GGGGSGSG 55  
 Db 149 GGGGSGSG 156  
 RESULT 10  
 ID O64634 PRELIMINARY; PRT; 2946 AA.  
 AC O64634;  
 DT 01-AUG-1998 (TRENBLrel. 07, Created)  
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CDC4 LIKE PROTEIN.  
 GN F17K2.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RC ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,  
 RA BRANSON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC003680; AAC06163.1; -  
 DR HSP; P04002; LWFA.  
 DR PFAM; PF00400; WD40; 1.  
 SQ SEQUENCE 2946 AA; 321930 MW; E6094427 CRC32;  
 Query Match 3.0%; Score 8; DB 10; Length 2946;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 LGGGSGSG 54  
 Db 1507 LGGGSGSG 1514  
 RESULT 11  
 ID Q43522 PRELIMINARY; PRT; 207 AA.  
 AC Q43522;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)

DE TFMS GENE.  
 GN TFMS.  
 OS Lycopodium esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
 OC Solanum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. UC82B; TISSUE=FRUIT;  
 RX MEDLINE; 97201476.  
 RA SANTINO C.G., STANFORD G.L., CONNER T.W.;  
 RT "Developmental and transgenic analysis of two tomato fruit enhanced  
 genes";  
 RL Plant Mol. Biol. 33:405-416(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. UC82B; TISSUE=FRUIT;  
 RA CONNOR T.W.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X95262; CAA64559.1; -  
 DR HSP; P24337; LHYP.  
 DR MENDEL; 16224; Lycos; 1531; 16224.  
 DR PFAM; PF00279; LTP; 1.  
 SQ SEQUENCE 207 AA; 18265 MW; B102F0EB CRC32;  
 Query Match 3.0%; Score 8; DB 10; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 GGGGSGSG 55  
 Db 45 GGGGSGSG 52  
 RESULT 12  
 ID O24568 PRELIMINARY; PRT; 256 AA.  
 AC O24568;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE GLYCINE-RICH PROTEIN PRECURSOR.  
 GN GRP3.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 OC Poaceae; Zea.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DK105; TISSUE=ROOT;  
 RX MEDLINE; 98187261.  
 RA GODDEMEIER M.L., WULFF D., FEIX G.;  
 RT "Root-specific expression of a Zea mays gene encoding a novel glycine-  
 rich protein, zmGRP3";  
 RL Plant Mol. Biol. 36:799-802(1998).  
 DR EMBL; Y07781; CAA69104.1; -  
 DR MENDEL; 27037; Zeama; 343; 27037.  
 KW Signal.  
 FT SIGNAL  
 SQ SEQUENCE 256 AA; 21712 MW; 0BF5ED59 CRC32;  
 Query Match 3.0%; Score 8; DB 10; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 6.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 GGGGSGSG 55  
 Db 242 GGGGSGSG 249

RESULT 13  
Q96463 PRELIMINARY; PRT; 288 AA.  
AC Q96463;  
DT 01-FEB-1997 (TRENBLrel. 02, Created)  
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE MYB4 TRANSCRIPTION FACTOR (FRAGMENT).  
GN MYB4.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
OC Poaceae; Hordeum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. HIMALAYA; TISSUE-ENDOSPERM;  
RA GUBLER F., ROBERTS J.K., JACOBSEN J.;  
RL Plant Physiol. 113:306-306(1997).  
EMBL; X99973; CAA68235.1; -;  
HSP; P01103; IPOM.  
MENDEL; 8597; Horvu; Myb; 8597.  
DR PFAM; PF00249; myb\_DNA-binding; 2.  
FT NON\_TER 1  
SQ SEQUENCE 288 AA; 31044 MW; FB45FD67 CRC32;  
  
Query Match 3.0%; Score 8; DB 10; Length 288;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 GGGGSGSG 55  
Db 265 GGGGSGSG 272  
|||||||  
  
RESULT 14  
Q89903 PRELIMINARY; PRT; 330 AA.  
AC Q89903;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE E1 PROTEIN.  
GN E1.  
OS Mouse cytomegalovirus 1.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Muromegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE; 90219186.  
RA BUEHLER B., KEIL G.M., WEILAND F., KOSZINOWSKI U.H.;  
RT "Characterization of the murine cytomegalovirus early transcription  
unit e1 that is induced by immediate-early proteins.";  
RL J. Virol. 64:1907-1919(1990).  
DR EMBL; L07320; AAA45907.1; -;  
DR EMBL; M35146; AAA45908.1; -;  
SQ SEQUENCE 330 AA; 34605 MW; E3FC3225 CRC32;  
  
Query Match 3.0%; Score 8; DB 12; Length 330;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 GGGGSGSG 55  
Db 238 GGGGSGSG 245  
|||||||  
  
RESULT 15  
O36415 PRELIMINARY; PRT; 252 AA.  
AC O36415;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE CAPSID PROTEIN.  
OS Alcelaphine herpesvirus 1 (wildbeest herpesvirus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C500;  
RX MEDLINE; 97404659.  
RA ENSSER A., PFLANZ R., FLECKENSTEIN B.;  
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";  
RL J. Virol. 71:6517-6525(1997).  
DR EMBL; AF003370; AAC58112.1; -;  
SQ SEQUENCE 252 AA; 24554 MW; 6B323798 CRC32;  
  
Query Match 3.0%; Score 8; DB 12; Length 252;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 GGGGSGSG 55  
Db 96 GGGGSGSG 103  
|||||||  
  
Search completed: February 11, 2000, 05:58:00  
Job time: 1059 sec





```
1057 ACAGGAGCTGCTGCCCTTTGCAATCTCAACAGCAGCACCAGCTTCGTG 1106
67 pProSerGlyProSerAlaProGluAlaValThrAla..... 79
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1107 GCGGGTTGGCATCTTCTCCATCTCGGCTTCATGCTCAGGAGCAGG 1156
80 .....ArgLeuVal.....GlyValLeuTrpPheValSer 89
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1157 CGTACCCATATCTGAGTGTGTAATCAGGCCCTGGCTGCATTCATCG 1206
90 ValThrThrGlyProTrp.....GlyAlaVa 98
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1207 CCTACCTCGAGCTGTGTGATGTTACCTTCTCGCCTTTGTGGCCTGC 1256
98 lAlaThrSerAla..GlyGlyGluGluSerLeuLys.....CysGlu 111
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1257 TGTTCCTTCATGTTGCTCTCGGACTAGACGCCATGCTTCATCG 1303
112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1304 .....GTGTGTAAGCCCTG..... 1321
128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysP 145
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1322 .....GACAGCGCTGGTGACATGATCC.... 1345
145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1346 .....CAGGTTGCTCCGTAAAGAACCGGAGGA....GATTCAT 1384
162 PheThrGlyAsnGluValGlyPhePheLysProLysSerCysArgAsnV 178
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1385 CCTCATCGTGTCTGCTCTCTTTCTTCATCGGCTCATTATGCTCACAG 1434
178 alAsnGly...TyrSerTyrLys..... 184
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1435 AGGCGGCATGATGCTGTGTCCAGCTCTTCGACTACTATGCGGCCAGTGGC 1484
185 .....ValAlaValAlaLeuSerLeuPheLeuGlyTriple 196
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1485 ATGTGCTCTCTTTGCGCCATCTTTCAGTCCCTCTGTGGCTGGT 1534
196 u...GlyAlaAspArgPheTyr.....LeuGlyTyrP 206
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1535 TTACGGAGCCAGCGCTCTATGACAACATGAAGATATGATGGGTACA 1584
206 roAlaLeuGlyLeuLysPheCysThrValGlyPheCysGlyIlegly 222
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1585 AGCGTGGCCCTTTATCAAACTGTTGGCTCTTTTTCACGCCAGCTGTG 1634
223 SerLeuIleAspPheIleLeuSerMetGlnIleValGlyProSerAs 239
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1635 TGCTGGCAACTCTCTGTTCTCCCTGATCAAAATACGCCACTACCTA 1684
239 pGlySerSerTyrIleLeuAspTyrTyrGlyThrArgLeu 252
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1685 CAACAAGATACATATCATGCTGGGGGATGCCCTG 1724
seq_name: /cgn2_5/ptodata/2/ina/PCUS9_COMB.seq: PCT-US93-01959-1
seq documentation_block:
; Sequence 1, Application PC/TUS9301959
; GENERAL INFORMATION:
; APPLICANT: Smith, E. Kelli
; APPLICANT: Borden, A. Laurence
; APPLICANT: Hartig, R. Paul
; APPLICANT: Weinsbank, L. Richard
; TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
```

```
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/01959
FILING DATE: 19930304
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40558A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2028 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
LIBRARY: rat brain
CLONE: rBl4b
FEATURE:
NAME/KEY: CDS
LOCATION: 126..1932
OTHER INFORMATION:
PCT-US93-01959-1
alignment_scores:
Quality: 88.50 Length: 248
Ratio: 0.763 Gaps: 15
Percent Similarity: 46.774 Percent Identity: 21.774
alignment_block:
US-09-060-609-2 x PCT-US93-01959-1
Align seg 1/1 to: PCT-US93-01959-1 from: 1 to: 2028
52 SerGlySerGlyGluLysValSerValSerLysMetAla...AlaAlaTr 67
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1057 ACAGGAGCTGCTGCCCTTTGCAATCTCAACAGCAGCACCAGCTTCGTG 1106
67 pProSerGlyProSerAlaProGluAlaValThrAla..... 79
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1107 GCGGGTTGGCATCTTCTCCATCTCGGCTTCATGCTCAGGAGCAGG 1156
80 .....ArgLeuVal.....GlyValLeuTrpPheValSer 89
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1157 CGTACCCATATCTGAGTGTGTAATCAGGCCCTGGCTGCATTCATCG 1206
90 ValThrThrGlyProTrp.....GlyAlaVa 98
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1207 CCTACCTCGAGCTGTGTGATGTTACCTTCTCGCCTTTGTGGCCTGC 1256
98 lAlaThrSerAla..GlyGlyGluGluSerLeuLys.....CysGlu 111
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1257 TGTTCCTTCATGTTGCTCTCGGACTAGACGCCATGCTTCATCG 1303
112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
```

```
1304 .....GTGTGTAGAAAGCTCGT..... 1321
128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysp 145
1322 .....GACAGCGCTGTGGACATGTATCC..... 1345
145 heProAlaProAsnIleThrCysLysaspSerSerGlyAsnGluThrHis 161
1346 .....CGGGTGTTCGGTAAAGAACCGGAGGA...GATTCTCAT 1384
162 .PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnV 178
1385 CTTCAATGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1434
178 alAsnGly...TyrSerTyrLys..... 184
1435 AGGCGCGCATGTAGCTGTCTCCAGCTCTTCGACTACTATGGCGCCAGTGGC 1484
185 .....ValAlaValAlaLeuSerLeuPheLeuGlyTyrLe 196
1485 ATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1534
196 u...GlyAlaAspArgPheTyr.....LeuGlyTyrP 206
1535 TTACGGAGCGCAGCGCTCTCTATGACAAATGAAGATATGATGGGTACA 1584
206 roAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCysGlyIleGly 222
1585 AGCGTGGCTCTTATCAATACTGTGGCTCTTTTCAGCCAGCTGTG 1634
223 SerLeuLeaAspPheIleLeuSerMetGlnIleValGlyProSerAs 239
1635 TGCGTGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1684
239 pGlySerSerTyrIleLeuAspTyrTyrGlyThrArgLeu 252
1685 CAACAAGAAGTACACATCATCATGGTGGGGGATGCCCTG 1724
seq_name: /cgn2_6/ptodata/2/ina/5d_comb.seq:us-08-700-013B-7
seq_documentation_block:
; Sequence 7, Application US/08700013B
; Patent No. 5919653
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian R.
; APPLICANT: Kowalski, Leslie R.Z.
; APPLICANT: Borden, Laurence A.
; APPLICANT: McKelvy, Jeffrey F.
; TITLE OF INVENTION: Human Glycine Transporter
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,013B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-108
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1597 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-700-013B-7

alignment_scores:
Quality: 87.00 Length: 302
Ratio: 0.617 Gaps: 17
Percent Similarity: 46.689 Percent Identity: 22.517

alignment_block:
US-09-060-609-2 x US-08-700-013B-7
Align seg 1/1 to: US-08-700-013B-7 from: 1 to: 1597

25 ThrGlyLeuTyrProMetArgGlyProPheLysAsnLeuAla.....Le 39
553 ACTGGACTTCATCTCTCCATGTGGGTAGCCAGTGGGCTGGGCAATG 602
39 uLeuProPheSerLeuProLeuLeuGlyGlyGly..... 51
603 TCTGGAGGTTTCCCTACCTGGCCTTCCAGAACGGGGAGGTGCTTTCCTC 652
52 .....SerGlySerGly.Glu..... 56
553 ATCCCTTACCTGATGATGCTGGCTCTGGCTGATACCCATCTCTCTCT 702
57 .LysValSerValSerLysMetAlaAlaIlePro..... 68
703 GGAGGTGTGGTGGCCAGTTTGCCAGCCAGGACGAGTGTGTGTGGA 752
69 .....SerGlyProSerAlaProGluAlaValThrAla 79
753 AGGCCATCCCGACTCTACAGGCTGTGGCATCGGATGCTGATCAACTCT 802
80 ArgLeuValGlyValLeuTyr..... 86
803 GTCCTAATAGCCATATACATGATGATTTGCTATACACTTTTCTA 852
87 .....PheValSerValThrGlyProTyrGlyAlaVala 99
853 CCTGTTTGCCTCCTTTGTGTGTACTA.....CCTGGGGCTCCTGCA 896
99 laThrSerAlaGlyGlyGluGluSerLeuLysCysGluAsp..... 112
897 ACAACCTCTTGAATACGCCAGAA.....TGCAAGATAAAACCAAA 937
113 LeuLysValGlyGlnTyrIleCysLysasp...ProLysIleAsnAspAl 128
938 CTTTATTAGATTCCTGTGTATCAGTGACCATCCCAAAATA..... 979
128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysp 145
980 .....CAGATCAAGAACTCGACTTTCGATGACCGGTT 1013
145 heProAlaProAsnIleThrCysLysaspSerSerGlyAsnGluThrHis 161
1014 ATCCCAACGTCACAATGTTAATTTACCAGCCAGGCCAATAAGACATT 1063
162 PheThrGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVa 178
1064 GTCAGTGGAGGTGAG...GAGTACTTCAAGTACTTTGTGCTGAAGATTTC 1110
178 lasnGlyTyrSerTyr.....LysValaValaLeuSerL 191
```







1422 TTTC 1425

seq\_name: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:US-08-414-657D-3

```
seq_documentation_block:
; Sequence 3, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 56...1069
; OTHER INFORMATION:
US-08-414-657D-3
```

```
alignment_scores:
Quality: 85.00 Length: 130
Ratio: 1.308 Gaps: 7
Percent Similarity: 50.000 Percent Identity: 27.692

alignment_block:
US-09-060-609-2 x US-08-414-657D-3 ..
Align seg 1/1 to: US-08-414-657D-3 from: 1 to: 1238

101 SerAlaGlyGluGluSerLeuLysCysGluAspLeuLysValGly.. 116
: : : : : : : : : : : : : : : : : : : : : : : : : : :
743 ACCAGGACGACAAAGCTTCCTCAATGTGAAGCTCAGCGGTGCCTGC 792
: : : : : : : : : : : : : : : : : : : : : : : : : : :

117 .....GlnTyrIleCysLysAspProLysIleAsnAspAlaThrG 130
: : : : : : : : : : : : : : : : : : : : : : : : : : :
793 ACCTGACTTTGAGTGGTACCGGGATGACACCGAGGATAAACAGTCAACG 842
```

```
130 lnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysPhePro 146
: : : : : : : : : : : : : : : : : : : : : : : : : : :
843 GCCTTGAGATTAGAGCACTGAG.....GCCAGTCTCTCCCTGACG 883
: : : : : : : : : : : : : : : : : : : : : : : : : : :
147 AlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHisPheTh 163
: : : : : : : : : : : : : : : : : : : : : : : : : : :
884 GTGACCAACGTCACT...GAGGAACACTACGGCAACTATACCTGTGTGGC 930
: : : : : : : : : : : : : : : : : : : : : : : : : : :
163 rGlyAsnGluValGly.....PhePheLysProI 173
: : : : : : : : : : : : : : : : : : : : : : : : : : :
931 TGCCAAACAGCTGGCGTCCACCAATGCCAGCTAGTCTCTTTTCAGACCG 980
: : : : : : : : : : : : : : : : : : : : : : : : : : :
173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
: : : : : : : : : : : : : : : : : : : : : : : : : : :
981 GTCGCGTGAGAGAAATCAACGGA...TCCATCAGTCTGGCGGTACCACTG 1027
: : : : : : : : : : : : : : : : : : : : : : : : : : :
190 SerLeuPheLeuGlyTrpLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1028 .....TGGCTG.....1033

206 oAlaLeuGlyLeuLeuLysPheCysThrValGlyPheCys 219
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1034 ....CTGGCAGCGTCCCTGTCTGCTTCTCAGCAAAATGT 1069
: : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-414-657D-5

seq_documentation_block:
; Sequence 5, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1014 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1014
```





```

159 GluThr.....HisPheThrGlyAsnGluValGlyPhePheLeuGlySProI 173
    |||
3586 GAAAAAATTAGAAATGATTTTGGCTATTCTAGTAAAGAAAAATGCTGT 3635
161 .....HisPheThrGlyAsnGluValGlyPhePheLeuGlySProI 173
3636 TTTCCAAGATGAATGTTATACATCCTATTCTGTAAATTTTTTGAAAAAAG 3685
173 leSerCysArgAsnValAsnGlyTyrSerTyrLysValAlaValAlaLeu 189
    ::::|
3686 TTCATGT..... 3693
190 SerLeuPheLeuGlyTyrPheLeuGlyAlaAspArgPheTyrLeuGlyTyrPr 206
    ::::|
3694 TCAGTTTTCCTAGT.....TTTACCTGTTTTC.. 3723
206 aAlaLeuLeuLeuLysPheCysThrValGlyPheCysGlyIleGlyS 223
    ::::|
3724 TCATAGGTCAGTAT...TCTGTGAAGCAAAAGATGCCTTTACCAT 3769
223 erLeuLeuAspPheIleLeuIleSerMetGluIleValGlyProSerAsp 239
    ::::|
3770 GAATCTTGAGTTTACATCAATAATTCGTATATAAGGGGATCAGAAGT 3819
240 GlySerSerTyrIleIleAspTyrTyrGlyThrArgLeuThrArgLeuSe 256
    ::::|
3820 AGGAGGAAAAAATAAGATAGACGAGAGAAAAAGAAAAACATTTTCCTCT 3869
256 rIleThrAsnGluThrPheArgLysThrGlnLeu 267
    |||||
3870 TATAACTTCTGAAGTAATTTGTAAAGAGATTG 3903

seq_name: /cgn2_6/ptodata/2/ina/5C_COMB.seq:US-08-509-187D-2

seq_documentation_block:
/ Sequence 2, Application US/08509187D
/ Patent No. 5834283
/ GENERAL INFORMATION:
/ APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
/ TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltrans-
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: LAHIVE & COCKFIELD, LLP
/ STREET: 28 State Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/509,187D
/ FILING DATE: 31-JUL-1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lampert Hammitte, Ann
/ REGISTRATION NUMBER: 34,858
/ REFERENCE/DOCKET NUMBER: DCI-033cpdv
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)227-7400
/ TELEFAX: (617)742-4214
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4079 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single

```





3770 GAATCTTGAGTTTACATCAATAATATGTATATTAAAGGGATCAGAAAGT 3819  
; Sequence 1, Application US/08276967  
240 GlySerSertyrilleAsPtyrGlyThrArgLeuThrArgLeuSe 256  
; : : : : :  
3820 AGGAAGGAAAAATAAGAGATAGCAGAGAAAAAGAAAAACATTCCTCT 3869  
256 rleThrAsnGluThrPheArgLysThrGlnLeu 267  
; : : : : :  
3870 TATAACTTCTGAAGTAATTTGTAAGAAAGATTG 3903  
seq\_name: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:US-08-276-967-1

seq\_documentation\_block:  
; Sequence 1, Application US/08276967  
; Patent No. 5851817  
; GENERAL INFORMATION:  
; APPLICANT: Hardy, Daniel M.  
; APPLICANT: Garbers, David L.  
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
; TITLE OF INVENTION: Sperm  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,967  
; FILING DATE: Submitted Herewith  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: UTSD:418\KIT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7785 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-276-967-1

alignment\_scores:  
Quality: 83.00 Length: 278  
Ratio: 0.709 Gaps: 16  
Percent Similarity: 42.086 Percent Identity: 23.022

alignment\_block:  
US-09-060-609-2 x US-08-276-967-1 ..  
Align seq 1/1 to: US-08-276-967-1 from: 1 to: 7785  
13 ArgAlaHisGlyGlnLysAsnThrArgArgAspGlyThrGlyLeuTyPr 29  
|||||  
2360 CGAGCTGCCGCCCAATGATCCCACTTTGAACCTGCGCTGCCCAAGTCC 2409  
29 oMetArgGlyProPheLysAsnLeuAla..... 38  
| |||||  
2410 TGCCAGAGCCCCACACCACTGTGAGCTTCTGCAAGCCGGCTGTGT 2459

39 .....LeuLeuProPhe.....SerLeu 44  
; : : : : :  
2460 CTGTGATCTTGCTTTTATTATCAGTGGCTCCCACTGCTCAACGGCTCTT 2509  
45 ProLeuLeuGlyGlyGlySerGlySerGlyGlyGlyLysValSerValSe 61  
; : : : : :  
2510 CCGTGTGATT.....GCTTCTACAAC 2529  
61 rLysMetaAlaAlaIatPrProSerGlyProSerAlaProGluAlaValT 78  
; : : : : :  
2530 GACAATTACTATAAGCTGGGACAGATTGTTTCAGCCCCA.....A 2570  
78 hrAlaArgLeuValGlyValLeuThrPheValSerValThrThrGlyPro 94  
; : : : : :  
2571 CTGCACAGAACATT.....GCCACTGCGGCCCA 2599  
95 .....TipGlyAlaValAlaThrSerAlaGlyGlyGluGluSerLe 108  
; : : : : :  
2600 GCAGTCGGATGGAGTGCCAGACCTTCAAGTGGCGGACACACACAGTG... 2646  
108 uLysCysGluAspLeuLysValGlyGlnTyrIleCysLys..... 121  
; : : : : :  
2647 .....TGCCAG...CTGAAGATGGCCAGTACGGCTGCCACCCCTATGGCA 2689  
122 .....AspPro..... 123  
2690 GTGCCACCTGCTCTGTCTACGGAGACCTCTCACTACCTCTCGACGGG 2739  
124 .....LysIleAsnAspAlaThrGlnGluProVa 133  
2740 AGGCGCTTTAACTTCATGGGCAAGTGCACCTACATCTTGGCCCCAACCC... 2787  
133 lAsnCysThrAsnTyrThrAlaHis..... 141  
2788 .....ValSerCysPheProAlaProAsnIleThr 151  
142 .....  
2834 AGGAGCGAGGACAGGAGGCGGTCTCTGCTTAAGCAAGGTCTACGTGACT 2883  
152 CysLysAspSerSer.....GlyAsnGluThrHisPheTh 163  
2884 CTGCCTGAAGCAGCACCGCTCACTCTGCTCAAGGGCAGACACACACGCTG 2933  
163 rGlyAsnGluValGlyPhePheLysProIleSerCysArgAsnVal.... 178  
2934 AGGTACGGAGTC...ACCTCCAGCCATACCTTCTAGAGGTGCTTCC 2980  
179 .....AsnGlyTyrSerTyrLysValAlaValAlaLeuSerLeuPhe 192  
2981 TGGCTCCCACTGGCGGATTGTGGAGCTGCACAGCGGTTCGGTCTCGCG 3030  
193 LeuGlyTyrLeuGlyAlaAspArgPheTyrLeuGlyTyrProAlaLeuGl 209  
; : : : : :  
3031 GTGAGATGGAGTGTGACACAGCAGCTGTTGTGAGTGTGCCAGCACCTT 3080  
209 yLeuLeuLysPheCysThrValGlyPheCysGly 220  
; : : : : :  
3081 CTCTGGCAAACTCTGT.....GGTCTCTGTGGC 3108  
seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-828-832-2

seq\_documentation\_block:  
; Sequence 2, Application US/08828832  
; Patent No. 5827711  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: NOVEL SUCCINATE DEHYDROGENASE SUBUNIT  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive

```

; CITY: Palo Alto
; STATE: CA USA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,832
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0250 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: 2454416
; US-08-828-832-2

```

```

alignment_scores:
  Quality: 82.00      Length: 193
  Ratio: 0.845       Gaps: 11
Percent Similarity: 50.259 Percent Identity: 24.870

alignment_block:
US-09-060-609-2 x US-08-828-832-2 ..

Align seg 1/1 to: US-08-828-832-2 from: 1 to: 649

82 ValGlyValLeuTrpPheValSerValThrGlyProTrpGlyAlaVa 98
   ::::::::::::::::::::
66 ATGGCGGTTCTCTGGAGGCTGAGT.....GCCGT 94

98 lAlaThrSerAlaGlyGly Glu.....GluSerLeuLysCysGlu 111
   ::::::::::::::::::::
95 TTGGGTGCCCTAGGAGCGGAGCTGTGTGCTTCCGAACCTCCAGTGTGCA 144

112 AspLeuLysValGlyGlnTyrIleCysLysAspProLysIleAsnAspAl 128
   ::::::::::::::::::::
145 GACCTGCTCATATCTCAGCATTTCTTCAGG.....ACCGACCT 182

128 aThrGlnGluProValAsnCysThrAsnTyrThrAlaHisValSerCysp 145
   ::::::::::::::::::::
183 ATCCAGCAATGTGTGGAGTGCAGCACATACACTGTGCACCGACCCACCA 232

145 heProAlaProAsnIleThrCysLysAspSerSerGlyAsnGluThrHis 161
   ::::::::::::::::::::
233 TTCTGGCTCCAGGCTGCA.....TCTCTCCAC 260

162 PheThrGlyAsnGluVal..... 167
   ::::::::::::::::::::
261 TGGACTAGCGAGAGGGTGTTCAGTGTGTTGCTCTGGGTCTGCTCCGGC 310

168 .GlyPhePheLysProIleSerCysArgAsnValAsnGlyTyrSerTyrL 184
   ::::::::::::::::::::
311 TGCTTATTGAATCCCTGCTCTCGCATGGAC.....TATT 345

```



---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2000, 07:47:33 ; Search time 13.25 Seconds  
(without alignments)  
957.626 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 269  
Sequence: 1 MHILKSPNVPRAHGKNT.....TRLRLSITNETFRKTLQYP 269

Scoring table: OLIGO

Searched: 142080 seqs, 47169319 residues

Database: PIR\_62:\*

Word size: 0

Number of hits that pass the threshold : 142080

1: pirl:\*  
2: pirl2:\*  
3: pirl3:\*  
4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.0	1520	1 TVFFA	protein-tyrosine k
2	8	3.0	358	2 S23383	protein kinase (EC
3	8	3.0	375	2 S22745	serine/threonine p
4	8	3.0	288	2 T05934	transcription fact
5	8	3.0	473	2 S70357	forhead transcrip
6	8	3.0	256	2 T03371	glycine-rich prote
7	8	3.0	252	2 T03160	capsid protein - a
8	8	3.0	427	2 H64938	hypothetical prote
9	8	3.0	529	2 S76187	hypothetical prote
10	8	3.0	2946	2 T00867	hypothetical prote
11	8	3.0	322	2 T04595	hypothetical prote
12	8	3.0	207	2 T07381	glycine-rich prote
13	8	3.0	548	2 T16082	hypothetical prote
14	8	3.0	173	2 A47303	hypothetical prote
15	8	3.0	1174	2 A40833	FTZ-F1 steroid rec
16	8	3.0	613	2 I39295	potassium channel
17	8	3.0	598	2 I38495	X-linked PEST-cont
18	7	2.6	419	1 JQ0346	ubiquinol-cytochr
19	7	2.6	451	1 JE0240	LIM kinase (EC 2.7
20	7	2.6	551	1 NRECE3	colicin E3 (EC 3.1
21	7	2.6	576	1 S22453	colicin E7 (EC 3.1
22	7	2.6	581	1 NDECE2	colicin E2 (EC 3.1
23	7	2.6	482	1 ZPBV	mitochondrial proc
24	7	2.6	577	1 A36442	mitochondrial proc
25	7	2.6	623	1 S33167	pointed protein, s
26	7	2.6	718	1 S33168	pointed protein, s
27	7	2.6	1959	1 AGRT	agrin - rat
28	7	2.6	166	1 KRBO2B	keratin, 68K type
29	7	2.6	183	1 KNRZG2	glycine-rich cell
30	7	2.6	835	1 A45596	trypanastigote-spe
31	7	2.6	511	1 IKECBB	colicin B - Escher
32	7	2.6	424	1 Z3BPBD	coat protein A pre
33	7	2.6	424	1 Z3BPF3	coat protein A pre
34	7	2.6	424	1 Z3BPF1	coat protein A pre
35	7	2.6	248	2 F71538	probable oxoacyl (

36 7 2.6 679 2 C71007 probable formate d  
37 7 2.6 399 2 I38901 JNK-activating pro  
38 7 2.6 638 2 I78846 LIM protein kinase  
39 7 2.6 617 2 I78847 LIM protein kinase  
40 7 2.6 638 2 JCS813 LIM-kinase (EC 2.7  
41 7 2.6 617 2 JCS814 LIM-kinase (EC 2.7  
42 7 2.6 163 2 S11399 PTS fructose-speci  
43 7 2.6 173 2 G71916 inorganic pyrophos  
44 7 2.6 173 2 D64597 inorganic pyrophos  
45 7 2.6 301 2 G72698 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

TVFFA  
protein-tyrosine kinase (EC 2.7.1.112) abl - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 15-Nov-1984 #sequence\_revision 30-Sep-1989 #text\_change 11-Jun-1999  
C:Accession: A28128; A00628  
R:Henkemeyer, M.J.; Bennett, R.L.; Gertler, F.B.; Hoffmann, F.M.  
Mol. Cell. Biol. 8: 843-853, 1988  
A:Title: DNA sequence, structure, and tyrosine kinase activity of the Drosophila mela  
A:Reference number: A28128; MUID:88174728  
A:Accession: A28128  
A:Molecule type: DNA  
A:Residues: 1-1520 <HEN>  
A:Cross-references: GB:M19692; GB:M18402; NID:gl58598; PIDN:AAA28934.1; PID:gl58600  
R:Hoffmann, F.M.; Fresco, L.D.; Hoffman-Falk, H.; Shilo, B.Z.  
Cell 35, 393-401, 1983  
A:Title: Nucleotide sequences of the Drosophila src and abl homologs: conservation an  
A:Reference number: A00628; MUID:84082064  
A:Accession: A00628  
A:Molecule type: DNA  
A:Residues: 'A', 375, 'AQ', 378-644, 'VGDV' <HOF>  
A:Cross-references: GB:K01042; NID:gl57175; PIDN:AAA28443.1; PID:gl57176  
C:Genetics:

A:Gene: abl  
A:Cross-references: FlyBase:FBgn0000017  
A:Introns: 112/1; 130/1; 310/2; 449/1; 644/3; 735/1; 775/2; 805/1; 1350/1  
C:Superfamily: Drosophila protein-tyrosine kinase abl; protein kinase homology; SH2 h  
C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; transformin  
F:211-260/Domain: SH3 homology <SH3>  
F:271-363/Domain: SH2 homology <SH2>  
F:386-645/Domain: protein kinase homology <KIN>  
F:394-402/Region: protein kinase ATP-binding motif  
F:417/Active site: Lys #status Predicted

Query Match 3.0%; Score 8; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGGSGG 55

Db 78 GGGGGSGG 85

##### RESULT 2

S23383  
protein kinase (EC 2.7.1.37) cdc2-related KKIALRE - human  
C:Species: Homo sapiens (man)  
C>Date: 15-Oct-1994 #sequence\_revision 15-Oct-1994 #text\_change 18-Jun-1999  
C:Accession: S23383; S22744  
R:Myerson, M.; Enders, G.H.; Wu, C.L.; Su, L.K.; Gorka, C.; Nelson, C.; Harlow, E.;  
EMBO J. 11, 2905-2917, 1992  
A:Title: A family of human cdc2-related protein kinases.  
A:Reference number: S23382; MUID:92347325  
A:Accession: S23383  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-358 <MEY>

A:Cross-references: EMBL:X66358; NID:g36614; PIDN:CAA47002.1; PID:g36615  
C:Superfamily: Kinase-related transforming protein; protein kinase homology  
C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase  
F:3-278/Domain: protein kinase homology <KIN>  
F:11-19/Region: protein kinase ATP-binding motif  
F:34,52,127,129/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 3.0%; Score 8; DB 2; Length 358;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 YPALGLLK 212  
|||||||  
DB 259 YPALGLLK 266

RESULT 3  
A:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 19-Dec-1997  
C:Accession: S22745  
R:Myerson, M.L.  
submitted to the EMBL Data Library, May 1992  
A:Reference number: S22743  
A:Accession: S22745  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-376 <MEY>  
A:Cross-references: EMBL:X66359  
C:Genetics:  
A:Introns: 152/3; 170/3  
C:Superfamily: Kinase-related transforming protein; protein kinase homology  
C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase  
F:3-296/Domain: protein kinase homology <KIN>  
F:11-19/Region: protein kinase ATP-binding motif

Query Match 3.0%; Score 8; DB 2; Length 376;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 YPALGLLK 212  
|||||||  
DB 277 YPALGLLK 284

RESULT 4  
A:Species: Hordeum vulgare (barley)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 16-Jul-1999  
C:Accession: T05954  
R:Gubler, F.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z15470  
A:Accession: T05954  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-288 <GUB>  
A:Cross-references: EMBL:X99973; NID:el251306; PIDN:CAA68235.1; PID:e258527  
A:Experimental source: cv. Himalaya, endosperm  
C:Genetics:  
A:Gene: myb4  
C:Superfamily: Arabidopsis 28K leaf-specific myb-related protein; myb DNA-binding repeat  
C:Keywords: DNA binding; transcription regulation  
F:60-110/Domain: myb DNA-binding repeat homology <MYB>

Query Match 3.0%; Score 8; DB 2; Length 288;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
|||||||  
DB 265 GGGGSGSG 272

RESULT 5  
S70357  
forkhead transcription factor HNF-3 alpha - human  
N:Alternate names: hepatocyte nuclear factor-3 alpha  
C:Species: Homo sapiens (man)  
C:Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 16-Jul-1999  
C:Accession: S70357  
R:Bingle, C.D.; Gowan, S.  
Biochim. Biophys. Acta 1307, 17-20, 1996  
A:Title: Molecular cloning of the forkhead transcription factor HNF-3-alpha from a hu  
A:Reference number: S70357; MUID:96254057  
A:Accession: S70357  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-473 <BIN>  
A:Cross-references: EMBL:U39840; NID:gl066121; PIDN:AA806493.1; PID:gl066122  
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology  
F:171-262/Domain: fork head DNA-binding domain homology <FHD>

Query Match 3.0%; Score 8; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
|||||||  
DB 277 GGGGSGSG 284

RESULT 6  
T03371  
glycine-rich protein grp3 - maize  
C:Species: Zea mays (maize)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 20-Aug-1999  
C:Accession: T03371  
R:Goddemeter, M.L.; Wulff, D.; Feix, G.  
Plant Mol. Biol. 36, 799-802, 1998  
A:Title: Root-specific expression of a Zea mays gene encoding a novel glycine-rich pr  
A:Reference number: Z14904; MUID:98187261  
A:Accession: T03371  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-256 <GOD>  
A:Cross-references: EMBL:Y07781; NID:el283638; PIDN:CAA69104.1; PID:e265498  
A:Experimental source: strain DK105; root  
C:Genetics:  
A:Gene: grp3  
C:Superfamily: glycine-rich cell wall structural protein 1

Query Match 3.0%; Score 8; DB 2; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55  
|||||||  
DB 242 GGGGSGSG 249

RESULT 7  
T03160  
capsid protein - alcelaphine herpesvirus 1  
C:Species: alcelaphine herpesvirus 1  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 07-May-1999  
C:Accession: T03160  
R:Essner, A.; Pfanz, R.; Fleckenstein, B.  
J. Virol. 71, 6517-6525, 1997

A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.

A:Reference number: Z14840; MUID:97404659

A:Accession: T03160

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-252 <ENS>

A:Cross-references: EMBL:AF005370; NID:92337967; PID:92338028

C:Keywords: capsid protein

Query Match 3.0%; Score 8; DB 2; Length 252;

Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

Db 96 GGGGSGSG 103

SLT 8

338

hypothetical protein b1784 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Nov-1997

C:Accession: H64938

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H64938

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-427 <BLAT>

A:Cross-references: GB:AE000273; GB:U00096; NID:g1788078; PID:g1788084; UWGP:b1784

A:Experimental source: strain K-12, substrain MG1655

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 427;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

Db 94 GGGGSGSG 101

SLT 9

167

hypothetical protein - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998

C:Accession: S76167

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S76167

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 <RAN>

A:Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PID:dl019159; PID:g1653513

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 529;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SLPLGGG 50

Db 260 SLPLGGG 267

RESULT 10

T00867

hypothetical protein F17K2.7 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 14-May-1999

C:Accession: T00867

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke

submitted to the EMBL Data Library, March 1998

A:Description: Arabidopsis thaliana chromosome II BAC F17K2 genomic sequence.

A:Reference number: Z14207

A:Accession: T00867

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2946 <ROU>

A:Cross-references: EMBL:AC003680; NID:g2979540; PID:g2979554

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Introns: 652/1; 913/3; 1165/3; 1210/3; 1245/3; 1283/3; 1309/2; 1336/3; 1438/3; 1638

; 2449/2; 2481/3; 2643/1; 2890/3; 2931/3

A:Note: F17K2.7

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 2946;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 LGGGSGSG 54

|||||

Db 1507 LGGGSGSG 1514

RESULT 11

T04595

hypothetical protein F23E13.150 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 14-May-1999

C:Accession: T04595

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse,

submitted to the Protein Sequence Database, March 1998

A:Reference number: Z15378

A:Accession: T04595

A:Molecule type: DNA

A:Residues: 1-322 <BEV>

A:Cross-references: EMBL:AL022141

A:Experimental source: cultivar Columbia; BAC clone F23E13

C:Genetics:

A:Map position: 4

A:Introns: 189/1

A:Note: F23E13.150

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 322;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

Db 149 GGGGSGSG 156

RESULT 12

T07381

glycine-rich protein Tfm5 - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 14-May-1999

C:Accession: T07381

R:Santino, C.G.; Stanford, G.L.; Conner, T.W.

Plant Mol. Biol. 33, 405-416, 1997

A:Title: Developmental and transgenic analysis of two tomato fruit enhanced genes.

A:Reference number: Z16000

A:Accession: T07381

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-207 <S>

A:Cross-references: EMBL:X95262; NID:gl166449; PID:e219171

A:Experimental source: cultivar UC82b; fruit

C:Genetics:

A:Gene: Tfm5

Query Match 3.0%; Score 8; DB 2; Length 207;

Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

45 GGGGSGSG 52

RESULT 13

T16082

hypothetical protein F16H11.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16082

R:Wu, X.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F16H11.

A:Reference number: Z18458

A:Accession: T16082

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-548 <W>

A:Cross-references: EMBL:U55376; NID:gl280130; PID:gl280133; PIDN:AAA98005.1; GSPDB:GN00

A:Experimental source: strain Bristol N2; clone F16H11

C:Genetics:

A:Gene: CESP:F16H11.5

A:Map position: 10

A:Introns: 71/1; 93/1; 187/3; 281/3; 316/3; 420/2; 460/1; 513/3; 545/3

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 548;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

153 KDSSGNET 160

|||||

109 KDSSGNET 116

RESULT 14

A47303

FTZ-F1 steroid receptor beta isoform - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 30-May-1997

C:Accession: A47303

R:Lavorgna, G.; Karim, F.D.; Thummel, C.S.; Wu, C.

Proc. Natl. Acad. Sci. U.S.A. 90, 3004-3008, 1993

A:Title: Potential role for a FTZ-F1 steroid receptor superfamily member in the control

A:Reference number: A47303; MUID:93219413

A:Accession: A47303

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-173 <LAV>

A:Note: sequence extracted from NCBI backbone (NCBIN:128795, NCBI:P:128796)

C:Genetics:

A:Gene: FlyBase:ftz-fl

A:Cross-references: FlyBase:FBgn0001078

C:Keywords: steroid hormone receptor

Query Match 3.0%; Score 8; DB 2; Length 173;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

Db 96 GGGGSGSG 103.

RESULT 15

A40853

potassium channel protein eag - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 24-Sep-1998

C:Accession: A40853

R:Warmke, J.; Drysdale, R.; Ganetzky, B.

Science 252, 1560-1562, 1991

A:Title: A distinct potassium channel polypeptide encoded by the Drosophila eag locus

A:Reference number: A40853; MUID:91262635

A:Accession: A40853

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1174 <WAB>

A:Cross-references: GB:M61157; NID:gl57311; PID:gl57312

C:Genetics:

A:Gene: FlyBase:eag

A:Cross-references: FlyBase:FBgn0000535

C:Keywords: transmembrane protein

Query Match

Best Local Similarity 3.0%; Score 8; DB 2; Length 1174;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGGGSGSG 55

|||||

Db 1125 GGGGSGSG 1132

Search completed: February 11, 2000, 08:02:52

Job time: 919 sec



OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC  
RX TISSUE-BRAIN:  
RA MEDLINE; 95220355.  
RA SCHRODER S., MORRIS S.A., KNORR R., PLESSMANN U., WEBER K.,  
RT VINH N.G., UNGEWICKELL E.;  
RT "Primary structure of the neuronal clathrin-associated protein  
RT auxilin and its expression in bacteria.";  
RL Eur. J. Biochem. 228:297-304(1995).  
CC -i- FUNCTION: BINDS CLATHRIN HEAVY CHAIN AND PROMOTES ITS ASSEMBLY  
CC INTO REGULAR CAGES.  
CC -i- TISSUE SPECIFICITY: BRAIN.  
CC -i- PTM: THE N-TERMINUS IS BLOCKED.  
CC -i- PTM: TARGET FOR COAT-ASSOCIATED CASEIN KINASE II IN VITRO.  
CC -i- SIMILARITY: CONTAINS A TENSIN DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U09237; AAA79037.1;  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; UNKNOWN\_1.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; UNKNOWN\_1.  
DR PROSITE; PS00076; DNAJ\_2; 1.  
KW Coat protein; Repeat; Phosphorylation.  
FT DOMAIN 33 44 3 X 4 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 33 36 1.  
FT REPEAT 37 40 2.  
FT REPEAT 41 44 3.  
FT DOMAIN 60 344 TENSIN.  
FT DOMAIN 406 414 SH3-BINDING (POTENTIAL).  
FT DOMAIN 463 757 PRO-RICH.  
FT DOMAIN 526 529 POLY-GLY.  
FT DOMAIN 846 910 DNAJ-LIKE.  
SQ SEQUENCE 910 AA; 99512 MW; CD5B90CA CRC32;

Query Match 5.5%; Score 79.5; DB 1; Length 910;  
Best Local Similarity 24.3%; Pred. No. 26;  
Matches 46; Conservative 22; Mismatches 86; Indels 35; Gaps 8;  
QY 33 PFKNLALLPFLPLGGGSGGSGEKVSKMAAW-PSGSPAPAVTAR----- 81  
Db 514 PPSNSELSS--DLFGGGAAGPVQSGSGVDVDFHFGPTSTQSTPERSATSTASPTL 570  
QY 81 LVGVLFVSVTTGTPWGAIVATSAGEESLKEDLVKGQYICKDKINDATQEPVNCNTYTA 140  
Db 571 RVG-----EGATFDPPFGAPSPSG-----QDL-LGSFLNTASASSDFFLOPTRSPSPV 618  
QY 141 HVSCFPAPNITCKDSSGNETHFTGNEVGFKPTSCRNWNGYSYKAV-----ALSLF-- 193  
Db 619 HASSTPAVNTQPDVSGAWDWHTRKPGFGMGSKSAATPTGSSGCTPTHQNKPTLDPPAD 678  
QY 193 LGWLGADRF 201  
Db 679 LGTLGGSSF 687

Search completed: February 11, 2000, 04:29:59  
Job time: 432 sec





```
QY 157 GNETHFTGNEVG-----FKPISCRVNGYSKVAVALSLFLGWLGADEFFYLGYPALG 209
Db 286 GNYTCAANKLGTNASLVLFPGSVRGING-SISLAVPL-----WL-----LAASLLC 333
QY 210 LLKFC 214
Db 334 LLSKC 338
RESULT 12
VG32_BPMD2
ID VG32_BPMD2 STANDARD; PRT; 223 AA.
AC O64226;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
GENE 32 PROTEIN (GP32).
32.
Mycobacteriophage D29.
OC Viruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98300335.
RA FORD M.E., SARKIS G.J., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
RL J. Mol. Biol. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022214; AAC18473.1;
SQ SEQUENCE 223 AA; 21822 MW; 6623D1E6 CRC32;
Query Match 5.6%; Score 80.5; DB 1; Length 223;
Best Local Similarity 30.7%; Pred. No. 4.3;
Matches 27; Conservative 8; Mismatches 36; Indels 17; Gaps 3;
QY 29 PMRGPPFNKALLPFSLP-----LLGGGSGSGEKVSVMKAAAWPSPSPAEAVTA 79
Db 37 PVLTPVTAVGAYTINIPAQAEFDIVLLGAGGGGQ-----MGSATAWGGGGGGSWTA 91
QY 80 RL---VGLWLFVSVTGPGWAVATSAGG 104
Db 92 TLRRGVDPNVAVTQITGVICGAGTAGPG 119
RESULT 13
ID IBP2_HUMAN STANDARD; PRT; 328 AA.
AC P18065; Q14619;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 2 PRECURSOR (IGFBP-2)
DE (IGFBP-2) (IGF-BINDING PROTEIN 2).
GN IGFBP2 OR BP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-RETINA;
RA AGARWAL N., HSIEH C.L., SILLS D., SWAROOP M., DESAI B., FRANCKE U.,
SWAROOP A.;
RT "Cloning, sequence analysis and expression of a cDNA encoding a novel
insulin-like growth factor binding protein (IGFBP-2).";
EMBO J. 8:2497-2502(1989).
[4]
SEQUENCE FROM N.A.
TISSUE-FETAL LIVER;
MEDLINE; 90060007.
RA BINKERT C., LANDWEHR J., MARY J.L., SCHWANDER J., HEINRICH G.;
RT "Cloning, sequence analysis and expression of a cDNA encoding a novel
insulin-like growth factor binding protein (IGFBP-2).";
EMBO J. 8:2497-2502(1989).
[3]
SEQUENCE FROM N.A.
TISSUE-FETAL LIVER;
MEDLINE; 90368661.
RA ZAPF J., KIEFER M., MERRYWEATHER J., MASIAZ F., BAUER D., BORN W.,
FISCHER J.A., FORESCH E.R.;
RT "Isolation from adult human serum of four insulin-like growth factor
(IGF) binding proteins and molecular cloning of one of them that is
increased by IGF I administration and in extrapancreatic tumor
hypoglycemia.";
J. Biol. Chem. 265:14892-14898(1990).
[2]
SEQUENCE FROM N.A. AND SEQUENCE OF 40-77.
MEDLINE; 90368661.
RA ZAPF J., KIEFER M., MERRYWEATHER J., MASIAZ F., BAUER D., BORN W.,
FISCHER J.A., FORESCH E.R.;
RT "Isolation from adult human serum of four insulin-like growth factor
(IGF) binding proteins and molecular cloning of one of them that is
increased by IGF I administration and in extrapancreatic tumor
hypoglycemia.";
J. Biol. Chem. 265:14892-14898(1990).
[1]
SEQUENCE FROM N.A.
TISSUE-PLACENTA;
MEDLINE; 91248211.
RA EHRENBORG E., VILHELMSDOTTIR S., BAJALICA S., LARSSON C., STERN I.,
KOCH J., BRONDUM-NIELSEN K., LUTHMAN H.;
RT "Structure and localization of the human insulin-like growth factor-
binding protein 2 gene.";
Biochem. Biophys. Res. Commun. 176:1250-1255(1991).
[5]
SEQUENCE FROM N.A.
TISSUE-PLACENTA;
MEDLINE; 92293319.
RA BINKERT C., MARGOT J.B., LANDWEHR J., HEINRICH G., SCHWANDER J.;
RT "Structure of the human insulin-like growth factor binding protein-2
gene.";
Mol. Endocrinol. 6:826-836(1992).
CC -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S37730; AAB23308.1;
DR EMBL; S37712; AAB22308.1; JOINED.
DR EMBL; S37722; AAB22308.1; JOINED.
DR EMBL; S37726; AAB22308.1; JOINED.
DR EMBL; M35410; AAA03246.1;
DR EMBL; X18302; CAA34373.1;
DR EMBL; M69241; AAA36048.1;
DR EMBL; M69237; AAA36048.1; JOINED.
DR EMBL; M69239; AAA36048.1; JOINED.
DR EMBL; M69240; AAA36048.1; JOINED.
DR EMBL; A09809; CAA00862.1;
DR PIR; S05389; S05389.
DR PIR; A35803; A35803.
DR PIR; JQ1029; JQ1029.
DR PIR; A41927; A41927.
DR PIR; A53957; A53957.
```

```
RT "Sequence analysis, expression and chromosomal localization of a
gene, isolated from a subtracted human retina cDNA library, that
encodes an insulin-like growth factor binding protein (IGFBP2).";
Exp. Eye Res. 52:549-561(1991).
[2]
SEQUENCE FROM N.A. AND SEQUENCE OF 40-77.
MEDLINE; 90368661.
RA ZAPF J., KIEFER M., MERRYWEATHER J., MASIAZ F., BAUER D., BORN W.,
FISCHER J.A., FORESCH E.R.;
RT "Isolation from adult human serum of four insulin-like growth factor
(IGF) binding proteins and molecular cloning of one of them that is
increased by IGF I administration and in extrapancreatic tumor
hypoglycemia.";
J. Biol. Chem. 265:14892-14898(1990).
[3]
SEQUENCE FROM N.A.
TISSUE-FETAL LIVER;
MEDLINE; 90060007.
RA BINKERT C., LANDWEHR J., MARY J.L., SCHWANDER J., HEINRICH G.;
RT "Cloning, sequence analysis and expression of a cDNA encoding a novel
insulin-like growth factor binding protein (IGFBP-2).";
EMBO J. 8:2497-2502(1989).
[4]
SEQUENCE FROM N.A.
TISSUE-PLACENTA;
MEDLINE; 91248211.
RA EHRENBORG E., VILHELMSDOTTIR S., BAJALICA S., LARSSON C., STERN I.,
KOCH J., BRONDUM-NIELSEN K., LUTHMAN H.;
RT "Structure and localization of the human insulin-like growth factor-
binding protein 2 gene.";
Biochem. Biophys. Res. Commun. 176:1250-1255(1991).
[5]
SEQUENCE FROM N.A.
TISSUE-PLACENTA;
MEDLINE; 92293319.
RA BINKERT C., MARGOT J.B., LANDWEHR J., HEINRICH G., SCHWANDER J.;
RT "Structure of the human insulin-like growth factor binding protein-2
gene.";
Mol. Endocrinol. 6:826-836(1992).
CC -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S37730; AAB23308.1;
DR EMBL; S37712; AAB22308.1; JOINED.
DR EMBL; S37722; AAB22308.1; JOINED.
DR EMBL; S37726; AAB22308.1; JOINED.
DR EMBL; M35410; AAA03246.1;
DR EMBL; X18302; CAA34373.1;
DR EMBL; M69241; AAA36048.1;
DR EMBL; M69237; AAA36048.1; JOINED.
DR EMBL; M69239; AAA36048.1; JOINED.
DR EMBL; M69240; AAA36048.1; JOINED.
DR EMBL; A09809; CAA00862.1;
DR PIR; S05389; S05389.
DR PIR; A35803; A35803.
DR PIR; JQ1029; JQ1029.
DR PIR; A41927; A41927.
DR PIR; A53957; A53957.
```

```
DR PIR: A00934; BBHU.
DR PIR: S14339; S14339.
DR PIR: S34075; S34075.
DR HSSP: P20231; 1AAO.
DR MIN: 138470; HUMAN.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PFAM: PF00084; sushi; 3.
DR PFAM: PF00089; trypsin; 1.
DR PFAM: PF00092; vwa; 1.
KW Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW Glycoprotein; MHC III; Repeat; Sushi; Signal; Polymorphism; Zymogen.
FT SIGNAL 1 25
FT CHAIN 26 764
FT CHAIN 26 259
FT CHAIN 260 764
FT DOMAIN 36 219
FT REPEAT 36 99
FT REPEAT 102 159
FT REPEAT 164 219
FT DOMAIN 482 764
FT ACT_SITE 526 526
FT ACT_SITE 576 576
FT ACT_SITE 699 699
FT DISULFID 37 76
FT DISULFID 62 98
FT DISULFID 103 145
FT DISULFID 131 158
FT DISULFID 165 205
FT DISULFID 191 218
FT CARBOHYD 122 122
FT CARBOHYD 142 142
FT CARBOHYD 285 285
FT CARBOHYD 378 378
FT CARBOHYD 291 291
FT VARIANT 28 28
FT VARIANT 28 28
FT VARIANT 32 32
FT VARIANT 736 736
FT CONFLICT 297 297
FT CONFLICT 300 300
FT CONFLICT 328 328
FT CONFLICT 356 357
FT CONFLICT 537 537
FT CONFLICT 764 764
SQ SEQUENCE 764 AA; 85532 MW; 711D8BD CRC32;

Query Match 5.8%; Score 83; DB 1; Length 764;
Best Local Similarity 24.1%; Pred. No. 11;
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

QY 24 GTGLPMRGPKNLALFFSLPLGG-----GSGSGEKVSV----- 61
D 2 GSNLSP-----QLCLMPFILGLLSGGVTTTPWSLARPGSGSLEGVKIGGSFRLAQEG 55
QY 61 SKMAAAMPSPG--PSAPEAVTARLVGLVFSVTTTGPWGAVALS-----AGGEESLKC-- 111
D 56 QALEVCPSPGPPVQTRCTR-----STGSWSTLTKDQKTVRKAECRAIHCP 105
QY 111 -EDLVGQYICKRKINDATQEPVNC--TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
D 106 PHDFENGWYWRSPYNNVSDISFHCYDGLRGSA-----NFTQVNGWSQTAICDNG 161
QY 167 VGFKF-----PISCRNVNGYSYKV 185
D 162 AGYCSNPGIPGTRKV--GSQYRL 183

RESULT 11
LAMP_HUMAN STANDARD; PRT; 338 AA.
AC Q13449;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LIMBIC SYSTEM-ASSOCIATED MEMBRANE PROTEIN PRECURSOR (LSAMP).
GN LAMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96235133.
RA PIMENTA A.F., FISCHER I., LEVITT P.;
RT "CDNA cloning and structural analysis of the human limbic-system-
associated membrane protein (LAMP).";
RL Gene 170:189-195(1996).
CC -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS
CC AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLLICULUS, SPINAL
CC CHORD AND CEREBELLUM.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE OBCAM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U41901; AAC50569.1;
CC MIM: 603241;
CC PFAM: PF00047; ig: 3.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW SIGNAL.
FT SIGNAL 1 28
FT CHAIN 29 315
FT PROPEP 316 338
FT DOMAIN 46 118
FT DOMAIN 146 204
FT DOMAIN 232 297
FT DISULFID 53 111
FT DISULFID 153 197
FT DISULFID 239 290
FT CARBOHYD 40 40
FT CARBOHYD 66 66
FT CARBOHYD 136 136
FT CARBOHYD 148 148
FT CARBOHYD 279 279
FT CARBOHYD 287 287
FT CARBOHYD 300 300
FT CARBOHYD 315 315
FT LIPID 315 315
SQ SEQUENCE 338 AA; 37308 MW; DF4014E4 CRC32;

Query Match 5.7%; Score 82.5; DB 1; Length 338;
Best Local Similarity 29.6%; Pred. No. 4.6;
Matches 37; Conservative 14; Mismatches 47; Indels 27; Gaps 7;

QY 101 SAGGEESLKCDLKVQ---QYICKDPKINDATQEPVNCNTYTAHVSCFPAPNITCKDSS 156
D 230 TTGRQASLKCEASVPAPDFEWRDTRNSANGLEIKSTE---GQSLTVTVNVT--EEHY 285
```

Db 272 PRAQCK--SDVFNQVKSISVGVNADTREALDFFSRGLVKAPIKILGSELASV 329  
QY 225 IDFILSMQIVG 236  
Db 330 YD-KMKVQIIVG 340

RESULT 10  
CFAB\_HUMAN STANDARD; PRT; 764 AA.  
AC P00751: O15006; Q29944;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE COMPLEMENT FACTOR B PRECURSOR (EC 3.4.21.47) (C3/C5 CONVERTASE)  
DE (PROPERDIN FACTOR B) (GLYCINE-RICH BETA GLYCOPROTEIN) (GBG) (PBF2).  
BF.  
Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RN SEQUENCE FROM N.A. (ALLELES S; FA AND FB).  
RX MEDLINE; 91065702.  
RA DAVRINCHE C., ABBAL M., CLERC A.;  
RT "Molecular characterization of human complement factor B subtypes.";  
RN Immunogenetics 32:309-312(1990).  
[2]  
RN SEQUENCE FROM N.A. (ALLELE S).  
RP TISSUE=LIVER;  
RC MEDLINE; 94237735.  
RX MEDLINE; 94237735.  
RA MEJIA J.E., JAHN I., DE LA SALLE H., HAUPTMANN G.;  
RT "Human factor B. Complete cDNA sequence of the BF\*S allele.";  
RN Hum. Immunol. 39:49-53(1994).  
[3]  
RN SEQUENCE FROM N.A. (ALLELE S).  
RP TISSUE=LIVER;  
RC MEDLINE; 94041399.  
RX SCHWABELE W., LUETTIG B., SOKOLOWSKI T., ESTALLER C., WEISS E.H.,  
RA MEYER ZUM BUESCHENFELDE K.-H., WHALEY K., DIPOLD W.;  
RT "Human complement factor B: functional properties of a recombinant  
RT zymogen of the alternative activation pathway convertase.";  
RN Immunobiology 188:221-232(1993).  
[4]  
RN SEQUENCE FROM N.A. (ALLELE S).  
RP MEDLINE; 94067177.  
RX HORIUCHI T., KIM S., MATSUMOTO M., WATANABE I., FUJITA S.,  
VOLANAKIS J.E.;  
RT "Human complement factor B: cDNA cloning, nucleotide sequencing,  
RT phenotypic conversion by site-directed mutagenesis and expression.";  
RN Mol. Immunol. 30:1587-1592(1993).  
[5]  
RN SEQUENCE FROM N.A.  
RP ROWEN L., DANKERS C., BASKIN D., FAUST J., LORETTZ C., AHEARN M.E.,  
RA BANTA A., SWARTZELL S., SMITH T.M., SPIES T., HOOD L.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
[6]  
RN SEQUENCE OF 26-764, PARTIAL SEQUENCE FROM N.A., AND CARBOHYDRATES.  
RX MEDLINE; 84161997.  
RA MOLE J.E., ANDERSON J.K., DAVISON E.A., WOODS D.E.;  
RT "Complete primary structure for the zymogen of human complement  
RT factor B.";  
RL J. Biol. Chem. 259:3407-3412(1984).  
[7]  
RN SEQUENCE OF 260-764.  
RX MEDLINE; 83204002.  
RA CHRISTIE D.L., GAGNON J.;  
RT "Amino acid sequence of the Bb fragment from complement Factor B.  
RT Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and  
RT completion of the sequence of the Bb fragment.";  
RN Biochem. J. 209:61-70(1983).  
[8]  
RN SEQUENCE OF 339-764 FROM N.A.  
RP

RX MEDLINE; 83273641.  
RA CAMPBELL R.D., PORTER R.R.;  
RT "Molecular cloning and characterization of the gene coding for human  
RT complement protein factor B.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:4464-4468(1983).  
[9]  
RN SEQUENCE OF 467-595 AND 752-764 FROM N.A.  
RX MEDLINE; 83039428.  
RA WOODS D.E., MARKHAM A.F., RICKER A.T., GOLDBERGER G., COITEN H.R.;  
RT "Isolation of cDNA clones for the human complement protein factor B,  
RT a class III major histocompatibility complex gene product.";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:5661-5665(1982).  
[10]  
RN SEQUENCE OF 16-259 FROM N.A.  
RX MEDLINE; 84158524.  
RA MORLEY B.J., CAMPBELL R.D.;  
RT "Internal homologues of the Ba fragment from human complement  
RT component factor B, a class III MHC antigen.";  
RL EMBO J. 3:153-157(1984).  
[11]  
RN SEQUENCE OF 1-99 FROM N.A.  
RP TISSUE=Blood;  
RC MEDLINE; 87102880.  
RA WU L.C., MORLEY B.J., CAMPBELL R.D.;  
RT "Cell-specific expression of the human complement protein factor B  
RT gene: evidence for the role of two distinct 5'-flanking elements.";  
RN Cell 48:331-342(1987).  
[12]  
RN GLYCATION IN POSITION 291.  
RP MEDLINE; 91174758.  
RA NIEMANN M.A., BROWN A.S., MILLER E.J.;  
RT "The principal site of glycation of human complement factor B.";  
RN Biochem. J. 274:473-480(1991).  
CC -1- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE  
CC COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND  
CC BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B  
CC TO GENERATE THE C3 OR C5 CONVERTASE. IT HAS ALSO BEEN IMPLICATED  
CC IN PROLIFERATION AND DIFFERENTIATION OF PREACTIVATED B  
CC LYMPHOCYTES, RAPID SPREADING OF PERIPHERAL BLOOD MONOCYTES,  
CC STIMULATION OF LYMPHOCYTE BLASTOGENESIS AND LYSIS OF ERYTHROCYTES.  
CC BA INHIBITS THE PROLIFERATION OF PREACTIVATED B LYMPHOCYTES.  
CC -1- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND  
CC C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH  
CC CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- POLYMORPHISM: TWO MAJOR VARIANTS, F AND S, AND 2 MINOR VARIANTS,  
CC AS WELL AS AT LEAST 14 VERY RARE VARIANTS, HAVE BEEN IDENTIFIED.  
CC -1- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX  
CC CLASS-III PROTEIN.  
CC -1- SIMILARITY: WITH COMPLEMENT C2.  
CC -1- SIMILARITY: CONTAINS 3 SUSHI (SCR) REPEATS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X72875; CAA51389.1; -  
DR EMBL; S67310; AAD13989.1; -  
DR EMBL; L15702; AAA16820.1; -  
DR EMBL; X00284; CAA25077.1; -  
DR EMBL; AF019413; AAB67977.1; -  
DR EMBL; K01566; CAB23472.1; -  
DR EMBL; J00125; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; J00126; AAA36226.1; -  
DR EMBL; J00185; AAA36219.1; ALT\_SEQ.  
DR EMBL; J00186; AAA36220.1; -  
DR EMBL; M15082; AAA59625.1; -

QY 210 LLKFTVGF 219  
DB 329 ASLCLLSKC 338

## RESULT 8

Y345\_MYCTU  
ID Y345\_MYCTU STANDARD; PRT; 487 AA.  
AC 006297;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL 52.2 KD TRANSPORT PROTEIN RV0346C.  
GN RV0346C OR MTCY13E10.06C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacterium.  
RN [1]  
SEQUENCE FROM N.A.  
STRAIN-H37RV;  
MEDLINE; 98295987.  
RA COLE S.T., BROCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,  
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,  
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,  
RA DAVIES R., DEVLIN K., FELTWELL T., GENTILES S., HAMLIN N., HOLROYD S.,  
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,  
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,  
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,  
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z95324; CAB08578.1;  
DR PROSITE: PS00218; AMINO-ACID\_PERMEASE; 1.  
DR PFAM: PF00324; aa\_permeases; 1.  
DB Hypothetical protein; Transport; Amino-acid transport; Transmembrane.  
FT TRANSMEM 26 46 POTENTIAL.  
FT TRANSMEM 50 70 POTENTIAL.  
FT TRANSMEM 98 118 POTENTIAL.  
FT TRANSMEM 133 153 POTENTIAL.  
FT TRANSMEM 163 183 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 256 276 POTENTIAL.  
FT TRANSMEM 290 310 POTENTIAL.  
FT TRANSMEM 341 361 POTENTIAL.  
FT TRANSMEM 369 389 POTENTIAL.  
FT TRANSMEM 414 434 POTENTIAL.  
FT TRANSMEM 440 460 POTENTIAL.  
SQ SEQUENCE 487 AA; 52194 MW; 64BBBCD CRC32;

Query Match 5.8%; Score 83.5; DB 1; Length 487;  
Best Local Similarity 26.5%; Pred. No. 5.7;  
Matches 27; Conservative 16; Mismatches 40; Indels 19; Gaps 5;

QY 159 EHFHTGNEVGFKPTSCRVNGYSKYKVAVALSLFLGLWLGADRFYLGYPALGLLKFTVGF 218  
DB 8 DERLTREDTGYHKLHSLRQLMIALGGAIGTGLFLG--AGGRLASAGPGL----FLVYGI 61

QY 219 CGIGSLIDFILISMQIVG-----PSDGS--SYIIDYYGTRL 252

DB 62 CGI-----FVFLIDRALGELVLRHPSGSGSFVSYAREFIGEKV 98

## RESULT 9

ADHL\_KLUMA  
ID ADHL\_KLUMA STANDARD; PRT; 348 AA.  
AC Q07288;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE ALCOHOL DEHYDROGENASE 1 (EC 1.1.1.1).  
GN ADH1.  
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Kluyveromyces.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-ATCC 12424;  
RX MEDLINE; 93250057.  
RA LADRIERE J.M., DELCOUR J., VANDENHAUTE J.;  
RT "Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase  
RT from Kluyveromyces marxianus ATCC 12424."  
RL Biochim. Biophys. Acta 1173:99-101 (1993).  
CC -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
CC -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.  
CC -!- SUBUNIT: HOMOTETRAMER.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X60224; CAA42785.1;  
DR PIR: S32521; S32521.  
DR PROSITE: PS00059; ADH\_ZINC; 1.  
DR PFAM: PF00107; adh\_zinc; 1.  
KW Oxidoreductase; Zinc; NAD; Multigene family.  
FT METAL 44 44 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 67 67 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).  
FT METAL 101 101 ZINC (SECOND ATOM) (BY SIMILARITY).  
FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).  
FT METAL 112 112 ZINC (SECOND ATOM) (BY SIMILARITY).  
FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 348 AA; 37158 MW; FA9C656F CRC32;

Query Match 5.8%; Score 83; DB 1; Length 348;  
Best Local Similarity 20.8%; Pred. No. 4.3;  
Matches 65; Conservative 38; Mismatches 107; Indels 102; Gaps 14;

QY 19 NTRRDG---TGLYPMRGPFKNLALLPFLPLGGGGSGSGEKVSKMAAWPSG----- 71

DB 37 NVKISGVCHTDLHAWQGDWP-----LDTKLPLV-GGHEGAGIVVANGENVVTGWEIGDYAGI 91

QY 71 -----PSAPEA-----VTARLV----- 83

DB 92 KWLNGSCMCECELSNEPNCRKADLSGYTHDGSFQQYATADAVQAARIPKNVDLAEVAP 151

QY 83 -----GV-----LWFSVTTGPGWAVATSAGGEESLKCDLKV----- 116

DB 152 ILCAGTVVYKALSAHIKAGDWVAISGACGLGSLAIQYAKAMGYRVLGIDAGDEKAKLF 211

QY 116 ----GOYICKPKINDATQEPVNCNTNTAHVSCFPANITCKDSGNETHTFTGNE--VGF 169

DB 212 KELGGEYFIDFTKDMVAEVIATNGVAHAVINVSVSEAAISTSVLYTRNGTWTVLVL 271

QY 170 FKPISCRNVNGYSKYKVAVALSLFLGLWG--AD-RFYLGYPALGLLK--FCIVGCGIGSL 224

```
Db 200 DAFYEHYQHGTFYAYYADKPSATTFLEFSVY-----IGDILTQYVYLPFCNPTAGSTFA 255
QY 245 IDYGTRLTR 254
Db 256 PRYWTPLVK 265
RESULT 4
EFIS_PORPU STANDARD; PRT; 515 AA.
AC EFIS_PORPU
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ELONGATION FACTOR 1-ALPHA S (EF-1-ALPHA S) (SPOROPHYTE-SPECIFIC EF-1-ALPHA).
GN TEF-S.
OS Porphyra purpurea.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
RN [1]
SEQUENCE FROM N.A.
RX STRAIN-AVONPORT;
RX MEDLINE; 96309386.
RA LIU O.Y., BALDAUF S.L., REITH M.E.;
RT "Elongation factor 1 alpha genes of the red alga Porphyra purpurea include a novel, developmentally specialized variant."
RL Plant Mol. Biol. 31:77-85(1996).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY IN THE SPOROPHYTE, A SHELL-BORING, FILAMENTOUS PHASE.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY. EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08841; AAA61790.1; -
DR HSP; Q01698; 1TU1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
Db Elongation factor; Protein biosynthesis; GTP-binding; Multigene family.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 91 95 GTP (BY SIMILARITY).
FT NP_BIND 151 154 GTP (BY SIMILARITY).
SQ SEQUENCE 515 AA; 56648 MW; 03F72B83 CRC32;
Query Match 6.2%; Score 89.5; DB 1; Length 515;
Best Local Similarity 24.7%; Pred. No. 1.9;
Matches 37; Conservative 20; Mismatches 56; Indels 37; Gaps 6;
QY 41 PFSLPL-----LGGGGSGSGEKVSYSKMAAAN-----PSGPSAPEAVTARLVGLWFSV 90
Db 261 PLRLPLQDVYKIGGTVPVGVGEVILKAGQVTFEPAGKAAVEKVSVM-----HH 313
QY 91 TTGPWGAATAGGGEESLKCEDLVKGOYICKDPK-----INDATQEPVNCN-- 138
Db 314 TSVQAIPEGDVNGVFNVLTKVDIKRGD-VCGDTKNDPPPTCEFLANVIIDHKNI RGY 372
QY 138 -----YTAHVSCFPAPNITCKDSSGNETH 161
Db 373 TPVLDCHTAHIAACKFASILSKDKRGKQTH 402
```

```
RESULT 5
LCT2_MOUSE STANDARD; PRT; 151 AA.
ID LCT2_MOUSE
AC O88803; O88804;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LEUKOCYTE CELL-DERIVED CHEMOTAXIN 2 PRECURSOR.
GN LECT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
SEQUENCE FROM N.A.
RX STRAIN-BALB/C; TISSUE-LIVER;
RX MEDLINE; 98382586.
RA YAMAGOE S., WATANABE T., MIZUNO S., SUZUKI K.;
RT "The mouse Lect2 gene: cloning of cDNA and genomic DNA, structural characterization and chromosomal localization."
RL Gene 216:171-178(1998).
CC -!- FUNCTION: HAS A NEUTROPHIL CHEMOTACTIC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE LECT2 / MIM-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB009687; BAA33383.1; -
DR EMBL; AB009688; BAA33384.1; -
DR MGD; MGI:1278342; LECT2.
KW Chemotaxis; Signal; Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 151 LEUKOCYTE CELL-DERIVED CHEMOTAXIN 2.
FT VARSPPLIC 98 151 FCVKIEYIPKPKYKSGIKKGEKGLTLLPLQKIYPIGQSHVH
FT VENCSDSDPTAYL -> QRLQAHHTLLNVFTCYWDKIQIPR
FT PTFRLCQNFLLH (IN ISOFORM LECT2Q).
SQ SEQUENCE 151 AA; 16405 MW; C5BD189A CRC32;
Query Match 6.0%; Score 86.5; DB 1; Length 151;
Best Local Similarity 24.8%; Pred. No. 0.83;
Matches 29; Conservative 9; Mismatches 52; Indels 27; Gaps 4;
QY 78 TARLVGLWFSVTTGPGAVATSGAGEESLKCEDLVKGOYICKDPKINDATQEPVNCN 137
Db 4 TTLLISALLSSALAGPWNACSKSSNEIRTCDSYCGGQYSAQ-----RTQR----- 52
QY 138 YTAHVSCFPAPNITCKDSSGNETHFTGNEVGFKPIKSNV-----NGSYKV 185
Db 52 ----H---HFGVDVLCDSGVVYAPFTGKIVGQEKPYRNKNAINDGIRLSGRGFCVKI 102
RESULT 6
ADH2_KLULA STANDARD; PRT; 348 AA.
ID ADH2_KLULA
AC P49383;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALCOHOL DEHYDROGENASE II (EC 1.1.1.1).
GN ADH2.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
RN [1]
SEQUENCE FROM N.A.
```

466 MGGTRAEDEVETN---VKCRVLPICIE---HGPRT-----FTKSTPCIIYNGHYFL 510

185 VAVALSFLGLWGLADRYFLYGPALGLKFKCTVTCGCGIGSLIDFLISMOIVGVPDGS 243

511 TTLYSIFLGVAVDRCLGYSAMAVGKMLTGLGFGIWIWDFLLVLVLGVLGPADDSW 569

RESULT 2

YJ05\_CAEEL STANDARD; PRT; 753 AA.

AC O01975;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL 85.1 KD NUCLEASE C41D11.5 (EC 3.1.30.-).

GN C41D11.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditis; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA GATTUNG S., MAGGI L.;

RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE DNA/RNA NON-SPECIFIC ENDONUCLEASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL; AF003740; AAC48141.1; .

DR WORMPEP; C41D11.5; CE08662.

DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.

DR PFAM; PF01223; Endonuclease; 1.

KW Hypothetical protein; Hydrolase; Nuclease; Endonuclease.

FT ACT\_SITE 593 593 BY SIMILARITY.

SC SEQUENCE 753 AA; 85068 MW; 9DF3803A CRC32;

Query Match 11.1%; Score 159.5; DB 1; Length 753;

Best Local Similarity 28.2%; Pred. No. 2.7e-06;

Hatches 46; Conservative 27; Mismatches 61; Indels 29; Gaps 5;

QY 104 GEESLKEE---DLKVGQYICKPKINDATOE---PVCNTNYTA-----HVSCEF 145

DB 284 GSAGLTCTFPDCDCRIGDTV---KVNCTSRKGCNPVSRNNVEAVCRFCWQLLPDGYDCE 339

QY 146 PAPNITCKDS-----SGNETHTGNEVFFKIPISCRNNGVSYKYVAVALSFLGLWGA 198

DB 340 PATNCTSTSKLLVTKCSAHSSVTCMGORNFYKRIPCNWSSGYSWTKMTILSVVLGGFGA 399

QY 199 DRYFLYGPALGLKFKCVTCGCGIGSLIDFLISMOIVGVPDGS 241

DB 400 DRYFLGLWKSNAIGLKFSEFGLGVWTLVDVVLIAVGYIKPIDGS 442

RESULT 3

VGL2\_CVMA5 STANDARD; PRT; 1324 AA.

AC P11224;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)

DE [CONTAINS: SPIKE PROTEIN S1 (90B); SPIKE PROTEIN S2 (90A)].

GN S.

OC Murine coronavirus MHV (strain A59).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	167.5	11.6	573	1	YK3_CAEEL	P34280 caenorhabdi	
2	159.5	11.1	753	1	XJ05_CAEEL	O01975 caenorhabdi	
3	90	6.3	1324	1	VGL2_CVMA5	PL1224 murine coro	
4	89.5	6.2	515	1	EF1S_PORPU	P50257 porphyra pu	
5	86.5	6.0	151	1	LC72_MOUSE	O88803 mus musculu	
6	85.5	5.9	348	1	ADH2_KLUJA	P43983 kluyveromyc	
7	85	5.9	338	1	LAMP_RAT	O62813 rattus norv	
8	83.5	5.8	487	1	Y346_MYCTU	O06297 mycobacteri	
9	83	5.8	348	1	ADH1_KLUJA	Q07288 kluyveromyc	
10	83	5.8	764	1	CFAB_HUMAN	P00751 homo sapien	
11	82.5	5.7	338	1	LAMP_HUMAN	Q13449 homo sapien	
12	80.5	5.6	223	1	VG32_BPMD2	O64226 mycobacteri	
13	80	5.6	328	1	IBP2_HUMAN	P18065 homo sapien	
14	80	5.6	410	1	LIPP_FIG	Q05991 sus scrofa	
15	79.5	5.5	910	1	AUX1_BOVIN	P27974 bos taurus	
16	79.5	5.5	4367	1	DYHC_NEUCR	P45443 neurospora	
17	79.5	5.5	855	1	POLG_HRV3	Q82081 human rhino	
18	79.5	5.5	3432	1	POLG_JAEV1	P27395 j genome po	
19	79.5	5.5	3432	1	POLG_JAEV5	P19110 j genome po	
20	79.5	5.5	3432	1	POLG_JAEVJ	P32886 j genome po	
21	79	5.5	468	1	LTP2_MOUSE	P17892 mus musculu	
22	79	5.5	525	1	STP_SPIOL	Q03411 spinacia ol	
23	78.5	5.5	245	1	CBS3_HORSE	Q19010 equus cabal	
24	78.5	5.5	398	1	MRL1_XENLA	P38565 xenopus lae	
25	78	5.4	557	1	YP85_MYCTU	Q50636 mycobacteri	
26	77.5	5.4	2179	1	POLG_HRV14	P03303 human rhino	
27	77	5.4	1239	1	NRG_DROME	P20241 drosophila	
28	76.5	5.3	772	1	MP1_YEAST	P35999 saccharomyc	
29	76	5.3	467	1	LIP1_HUMAN	P54315 homo sapien	
30	76	5.3	1556	1	VG67_HSV11	Q00107 ictaluriid h	
31	75.5	5.2	350	1	ADH1_KLUJA	P20369 kluyveromyc	
32	75.5	5.2	592	1	CEA_CITFR	P04480 citrobacter	
33	75.5	5.2	1114	1	DEF1_CHICK	P36197 gallus gall	
34	75.5	5.2	682	1	NISP_LACLA	Q07596 lactococcu	
35	75.5	5.2	484	1	PER2_VOLCA	P81132 volvox cart	
36	75.5	5.2	676	1	PER_DROMA	Q03353 drosophila	
37	75.5	5.2	1224	1	PER_DROME	P07663 drosophila	
38	75.5	5.2	661	1	PER_DROSE	Q03354 drosophila	
39	75.5	5.2	656	1	PER_DROSI	Q03355 drosophila	

90 VTTGPGAVATSAGGEESLKCDLKVGQYICKDP-----K 124  
 !:|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!  
 415 VSTNPLGPV-----VCELENSF ILCEDPVPLGPGQTGOQPANESFNECKCLK 465  
 !:|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!  
 125 INDATGEPVNCTNYTARVSCFPAPNTICKDSSGNETHFTGNEVGFPPKPISCRNVNGYSYK 184  
 !:|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!|!

submitted to the EMBL Data Library, August 1998

A:Description: The sequence of A. thaliana T24H24.

A:Reference number: Z14333

A:Accession: T01466

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-171 <OU>

A:Cross-references: EMBL:AF075598; NID:G3293581; PIDN:AAC28213.1; PID:G3377840

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 98/1; 116/2

A:Note: T24H24.11

C:Superfamily: Yeast nitrogen fixation protein: nitrogen fixation homology

F:61-113/Domain: nitrogen fixation protein homology <NFI>

Query Match	5.6%;	Score 80;	DB 2;	Length 171;
Best Local Similarity	26.5%;	Pred. No. 5.9;		
Matches 39;	Conservative 16;	Mismatches 62;	Indels 30;	Gaps 7;
8 PNVI-----PRAHG--QKNRRDGTGLY--PMRGPFKNLALLPFLSLPLGGGS-----53				
30 PNVIDHYDNPNNVSGSDKDPNVTGVLGAPQCGDVNKLQV-----KFDGSGQIIDAKF 83				
53 ---GSGGEKVSVKMAAAMPSPGSAPEAVTARLVGLVWFVSVTTGPGWAVATSAAGEESLK 109				
84 KIFGCGSAIAASSVATEWVKGSVEEVLTIKNSQIAKHLSL---PPVKLHCSMLAEADAIK 140				
110 CEDLKVGGQICKDPKINDATQEPVNCVT 136				
141 A----AIKNYKEKODKANGETVETIDST 164				

Search completed: February 11, 2000, 07:38:37  
Job time: 878 sec





A:Reference number: JC4776; MUID:96235133

A:Accession: JC4776

A:Molecule type: mRNA

A:Residues: 1-338 <PIM>

A:Cross-references: GB:U41901; NID:gl276899; PID:gl276899

A:Experimental source: brain

C:Comment: This protein is a neuronal surface glycoprotein distributed in cortical and subcortical regions. It plays development and function in limbic system.

C:Genetics:

A:Gene: lamp

C:Keywords: brain; glycoprotein; membrane protein; phosphoprotein

F:1-7/Domain: signal sequence #status predicted <SIG>

F:333-338/Region: hydrophobic

F:40-66,136-148,279,287,300,315/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:42,115,142,164,171,220,231/Binding site: phosphate (Thr) (covalent) #status predicted

F:95,192,204,236,310/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 5.7%; Score 82.5; DB 2; Length 338;

Best Local Similarity 29.6%; Pred. No. 7.8;

Matches 37; Conservative 14; Mismatches 47; Indels 27; Gaps 7;

QY 101 SAGGEESLACEDLVKVG-----QVICKDKINDATQEPVNCNTNTAHVSCFPAPNITCKDSS 156

DB 230 TTGRQASLACEASVPAPDFEWYRDTRINSANGLEIKSTE---GQSSLTVTNTVEHY 285

QY 157 GNETHFTGNEVG-----FFPISCRNVNGYSYKVAVALSLFLWGLGADRFYLGYPALG 209

DB 286 GNYTCVAANKLGVNTNASLVFRPGRSGVING-SISLAVPL-----WL-----LAASLLC 333

QY 210 LLKFC 214

DB 334 LLSKC 338

RESULT 11

T03788

laccase (EC 1.10.3.2) - common tobacco (fragment)

N:Alternate names: diphenol oxidase; urishiol oxidase

C:Species: Nicotiana glauca (common tobacco)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999

C:Accession: T03788

R:Kiefer-Meyer, M.; Gomord, V.; O'Connell, A.; Halpin, C.; Paye, L.

Gene 178, 205-207, 1996

A:Title: Cloning and sequence analysis of laccase-encoding cDNA clones from tobacco.

A:Reference number: JC5229; MUID:97080572

A:Accession: T03788

Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-409 <KIE>

A:Cross-references: EMBL:U45243; NID:gl685090; PIDN:AAC49538.1; PID:gl685091

C:Superfamily: laccase

C:Keywords: copper; oxidoreductase

Query Match

Best Local Similarity 5.7%; Score 82; DB 2; Length 409;

Matches 47; Conservative 24; Mismatches 73; Indels 64; Gaps 8;

QY 63 MAANP--SGPSAPEAVTAR-----LVGLVFTVSVTTGPGAVATSGAGBESLKCEDL 113

DB 132 MVAASFMDAPVADNITATLHYSGALGTSTPTLTSTPPQNATSVANNFLDALK--SL 189

QY 114 KVGQVICKDKINDATQEPVNCNTNTAHVSCFPAPNITCKDSSGNETHTGNEVGFFKPI 173

DB 190 NSKKYPAKVPQVVD-----HSLFTAGLGNPCP--TCKQANGRWASVNNVTVP-- 241.

QY 174 SCRNVNGYSYKVAVALSLFLWGLGADRFYLGYPALGLLKFCTVFGCGISGLDIFILISMQ 233

DB 241 -----TVALQAHFFGNGV--FTTDFPA----- 263

QY 234 IVGPSDGSYSIIYDXTGTRILSITNET 261

DB 241 -----TVALQAHFFGNGV--FTTDFPA----- 263

QY 234 IVGPSDGSYSIIYDXTGTRILSITNET 261

DB 241 -----TVALQAHFFGNGV--FTTDFPA----- 263

DB 263 -----NPPFVENVGTGPTNLTATNGT 284

RESULT 12

A54080

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 23-Jul-1999

C:Accession: A54080; I50592

R:Fang, K.S.; Barker, K.; Sudol, M.; Hanafusa, H.

J. Biol. Chem. 269, 14056-14063, 1994

A:Title: A transmembrane protein-tyrosine phosphatase contains spectrin-like repeats

A:Reference number: A54080; MUID:94245724

A:Accession: A54080

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1237 <FAN>

A:Cross-references: EMBL:Z21960; NID:G510510; PIDN:CAA79972.1; PID:G510511; GB:L13285

C:Superfamily: leukocyte common antigen; leukocyte common antigen cytosolic domain homolog

C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase

F:528-1170/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F:786/Active site: Cys (phosphocysteine intermediate) #status predicted

F:792/Binding site: substrate phosphate (Arg) #status predicted

Query Match 5.6%; Score 80.5; DB 2; Length 1237;

Best Local Similarity 21.6%; Pred. No. 50;

Matches 48; Conservative 30; Mismatches 75; Indels 69; Gaps 11;

QY 71 PSAPAEAVTA---RLVGLVLFVSVTTGPGAVATSGAG-----EESLACE 111

DB 323 PKAPENLVHTDRNVTWTW----KPTGTLEKHIDGYVECNNTSQNVNRNETSFTCG 377

QY 112 DL-----KVGQVICKD---PKINDATQ---EPVNCNTNTAHVSCFPAPNIT 151

DB 378 DLEPYSTGVSIVRAFKSKYKNNFEKEKVGSGFQTPAKPENVTDFKLLTADNTVKA 437

QY 152 CKDSS--GNETHF-----TGNEVGFFKPISCRNVN---GYSYKVAVALSLF 192

DB 438 CRQKVGNGKTEKFLKSNSSNGENQKNECF---TVRDLVLTQYTFKISVNGVY 493

QY 193 LGWLGADRFYLGYPALGLLKFCTVFGCGISGLDIFILISMQI 234

DB 494 TGDSCVEEIIYTRYNSRALIIFLV--FLIVVTSLALLLVLYKI 533

RESULT 13

LIPG

triacylglycerol lipase (EC 3.1.1.3) - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 30-Nov-1980 #sequence\_revision 22-May-1981 #text\_change 24-Feb-1995

C:Accession: A90638; A91256; A90675; A91124; A00732

R:De Caro, J.; Boudouard, M.; Bonicel, J.; Guidoni, A.; Desnuelle, P.; Ravery, M.

Biochim. Biophys. Acta 671, 129-138, 1981

A:Title: Porcine pancreatic lipase. Completion of the primary structure.

A:Reference number: A90638; MUID:82113655

A:Accession: A90638

A:Molecule type: protein

A:Residues: 308-449 <DEC>

A:Experimental source: pancreas

R:Blanchetta, J.D.; Bidaud, J.; Guidoni, A.A.; Bonicel, J.J.; Ravery, M.

Eur. J. Biochem. 97, 395-405, 1979

A:Title: Porcine pancreatic lipase. Sequence of the first 234 amino acids of the peptide

A:Reference number: A91256; MUID:79236335

A:Contents: carbohydrate-binding site

A:Accession: A91256

A:Molecule type: protein

A:Residues: 1-234 <BIA>

A:Experimental source: pancreas

R:Guidoni, A.; Bonicel, J.; Blanchetta, J.; Ravery, M.

Biochimie 61, 841-845, 1979

A:Title: Porcine pancreatic lipase. Sequence between the 235th and 307th amino acids.

A:Reference number: A90675; MUID:80088446

A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CAl>  
A:Cross-references: GB:J00126; NID:gl87723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: evidenc  
A:Reference number: A25971; MUID:87102880  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <WUL>  
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Bhow, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement Factor B.  
A:Reference number: S14339; MUID:91174758  
A:Accession: S14339  
A:Molecule type: Protein  
A:Residues: 270-329 <NTE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologues of the Ba fragment from human complement component factor B  
A:Reference number: A44628; MUID:84158524  
A:Accession: A44628  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225, 'F', 227-259 <MOR>  
R:Schwaible, W.; Luttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A:Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A:Reference number: I54409; MUID:94041399  
A:Accession: I54409  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: GB:S67310; NID:9452937  
R:Horluchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conv  
A:Reference number: I57824; MUID:94067177  
A:Accession: I57824  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-31, 'Q', 33-764 <RE2>  
A:Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C:Comment: 292-Cys has a free sulfhydryl.  
C:Genetics:  
A:Gene: GDB:BF  
A:Cross-references: GDB:119726; OMIM:138470  
A:Map position: 6p21.3-6p21.3  
A:Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 69  
A:Note: the list of introns may be incomplete  
A:Note: gene is located in the major histocompatibility complex, class III region  
C:Complex: complement factor B initially forms an inactive complex with complement facto  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha c  
al  
A:Pathway: complement alternate pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology  
C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydroly  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <WAT>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <FHI>  
F:103-158/Domain: complement factor H repeat homology <FH2>  
F:163-218/Domain: complement factor H repeat homology <FH3>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>

F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76,62-98,103-145,131-158,165-205,191-218,478-596,511-527,599-615,656-682,695-725  
F:122,142,285,378/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526,576,699/Active site: His, Asp, Ser #status experimental

Query Match 5.8%; Score 83; DB 1; Length 764;  
Best Local Similarity 24.1%; Pred. No. 18;  
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

QY 24 GTGLPMRGPFKNLALLPFLPLGG-----GSGSGEKVS----- 61  
DB 2 GSNLSP-----QLCLMFFLLGLSGGVTTTPWSLARPGSGSLEGVEIKGSPRLQEG 55

QY 61 SKMAAAWPSG-PSAPEAVTARLVGLWVSVTTTPGAVATS-----AGGESLKC-- 111  
DB 56 QALEYVCPSEFYFPVQTRCR-----STGSLTLKTQDKTVRKAECAHCPR 105

QY 111 -EDLVGQYICKDKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166  
DB 106 PHDFENGGEYWRSPYNNVSDSEIFHCYDGYTLRCSA----NRTQVNGRWSSQTAICDNG 161

QY 167 VGPFK-----PISCRNNGYSYKV 185  
DB 162 AGYCSNPGIPIGTRKV-GSQYRL 183

RESULT 9

T10729

C:Species: Dunaliella salina

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C:Accession: T10729

R:Fisher, M.; Gokhman, I.; Pick, U.; Zamir, A.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z17101

A:Accession: T10729

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1274 <FIS>

A:Cross-references: EMBL:U77059; NID:gl684791; PID:gl684792

C:Genetics:

A:Gene: ttf1

C:Superfamily: transferrin repeat homology

Query Match 5.8%; Score 83; DB 2; Length 1274;  
Best Local Similarity 22.9%; Pred. No. 31;  
Matches 30; Conservative 17; Mismatches 50; Indels 34; Gaps 6;

QY 57 KVSVSKMAAAWPSGSAPEAV-TARLVGLWVSVTTTPGAVATSAGGESLKCEDLKV 115  
DB 579 QVDAETIEKFWEDNVCAPGSTENGPLIC-----GKYGEVGENGG----- 620

QY 116 GQYICKDKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGEFKPISC 175  
DB 620 ---LCRKCKTDCSEDPY--AGYDGAHVCI-----DDDDGNQ--FTGGDIAPVYKHTL 665

QY 176 RNVNGYSYKVA 186  
DB 666 RYNGPNLNTA 676

RESULT 10

JC4776

C:Species: Homo sapiens (man)

C:Date: 10-May-1996 #sequence\_revision 16-Aug-1996 #text\_change 13-Nov-1998

C:Accession: JC4776

R:Pimenta, A.F.; Fischer, I.; Levitt, P.

Gene 170, 189-195, 1996

A:Title: cDNA cloning and structural analysis of the human limbic-system-associated m





S55661  
hypothetical protein 66 - equine herpesvirus 2  
C:Species: equine herpesvirus 2  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-May-1999  
C:Accession: S55661  
R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
J. Mol. Biol. 249, 520-528, 1995  
A:Title: The DNA sequence of equine herpesvirus 2.  
A:Reference number: S55594; MUID:95302501  
A:Accession: S55661  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-456 <TEL>  
A:Cross-references: GB:U20824; NID:9695172; PID:4695239  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February













PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 1; Page 456; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
CC Sequence 455 BP; 102 A; 107 C; 115 G; 122 T;

Query Match 52.5%; Score 425.2; DB 1; Length 455;  
Best Local Similarity 96.3%; Pred. No. 3.8e-119;  
Matches 439; Conservative 3; Mismatches 12; Indels 2; Gaps 1;  
QY 177 GGTCCTCAAGATGGCGCCCTGGCCGTCTGGTCCGTCTCCGGAGCCGTGACGGC 236  
DB 2 GGTCCTCAAGATGGCGCCCTGGCCGTCTGGTCCGTCTCCGGAGCCGTGACGGC 61  
QY 237 CAGACTCGTTGGTCTGTGGTTCGTCTCAGTCACTACAGGACCCCTGGGGGCTGTGC 296  
DB 62 CAGACTCGTTGGTCTGTGGTTCGTCTCAGTCACTACAGGACCCCTGGGGGCTGTGC 121  
QY 297 CACCTCCGGCGGGGCGAGGAGTCGTCTAGTCGGAGGACCTCAAAAGTGGACAATATAT 356  
DB 122 CACCTCCGGCGGGGCGAGGAGTCGTCTAGTCGGAGGACCTCAAAAGTGGACAATATAT 181  
QY 357 TTGTAAGATCCAAAATAAATGACGCTACGCAAGAACCACTGTTAACTGTACAACTACAC 416  
DB 182 TTGTAAGATCCAAAATAAATGACGCTACGCAAGAACCACTGTTAACTGTACAACTACAC 241  
QY 417 AGCTCATGTTCTCTGTTTCCAGCACCCCAACATACTGTGAAGATTCCAGTGGCAATGA 476  
DB 242 AGCTCATGTTCTCTGTTTCCAGCACCCCAACATACTGTGAAGATTCCAGTGGCAATGA 301  
QY 477 AACACATTTTACTGGAGCGAAGTGGTCTTTTCAAGCCCATATCTGCCGAATGTAA 536  
QY 302 AACACATTTTACTGGAGCGAAGTGGTCTTTTCAAGCCCATATCTGCCGAATGTAA 361  
QY 537 TGGCTATTCTCAAAATGGCAGTCGATGCTCTCTTTTCTTGGATGTTGGAGCAGA 596  
DB 362 TGGCTATTCTCTAC--NNTAGCAGTNNNTGCTCTTTTCTTGGATGTTGGAGCAGA 419  
QY 597 TCGATTTTACCTTGATACCCCTGCTTTGGTGTGTT 632  
DB 420 TCGATTTTACCTTGATACCCCTGCTTTGGTGTGTT 455

RESULT 4  
T74893  
ID T74893 standard; cDNA; 1951 BP.  
AC T74893;  
DT 02-OCT-1997 (first entry)  
DE Mouse neurogenic differentiation protein (NeuroD2) DNA clone 1.1.1.  
KW Neurogenic differentiation protein; NeuroD; neuroD2 gene;  
KW transcriptional activator; neuron; pancreas; gastrointestinal;  
KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;  
KW SS.  
OS Mus musculus.  
FT Key Location/Qualifiers  
FT cds 230..1378  
FT /\*tag= a

PN WO9716548-A1.  
PD 09-MAY-1997.  
PF 30-OCT-1996; U17532.  
PR 02-NOV-1995; US-552142.  
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PA (WEIN-) WEINTRAUB N.  
PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;  
DR WPI; 97-272117/24.  
DR P-PSDB; W22442.  
PT Nucleic acid encoding neurogenic differentiation polypeptide -  
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal  
PT development  
PT Claim 1; Page 69-70; 81pp; English.  
PS Neurogenic differentiation (NeuroD) genes (T74887-94) and proteins  
CC (W22436-43) from human, mouse and frog have been identified,  
CC isolated and sequenced. NeuroD polypeptides are tissue-specific  
CC basic-helix-loop-helix (bHLH) transcriptional activators involved  
CC in neuronal, endocrine and gastrointestinal development. They were  
CC discovered by expression cloning and screening assays designed to  
CC identify possible bHLH proteins capable of interacting with the  
CC protein product of the Drosophila daughterless gene. Novel neuroD2  
CC and neuroD3 genes, related to neuroD1, have been identified.  
CC NeuroD nucleic acids can be used to produce NeuroD polypeptides,  
CC construction of test cell lines, as probes, in gene therapy, and to  
CC produce transgenic animals as models of disease.  
CC Sequence 1951 BP; 359 A; 668 C; 557 G; 353 T;  
SQ

Query Match 4.5%; Score 36.4; DB 1; Length 1951;  
Best Local Similarity 49.0%; Pred. No. 0.31;  
Matches 97; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 121 CCCTCTCTCCCTCCCGCTCTCTGGCGGAGGCGGAGCGGAAGTGGCGAGAAAGTGTGCGTC 180  
DB 260 CTCCTCTCGGACGTGCCCAAGTTCGCCAGCTGGGGCGGACGACGACGAGCCGAGG 319  
QY 181 TCCAAGATGCGCGCGCGCTGCGCTCTGCTCCGTCTGCTCCGGAGCGCGTGACGGCCAGA 240  
DB 320 AGCGACAAGGCGGCG 379  
QY 241 CTGCTTGGTCTCTGTTGGTCTGCTCAGTCACTACAGGACCCCTGGGGGCTGTGCGCAC 300  
DB 380 CCG 439  
QY 301 TCCGCGCGGCGCGGAGGAG 318  
DB 440 TTGCTGAGGTCAAGGAG 457

RESULT 5  
V42934  
ID V42934 standard; cDNA; 1951 BP.  
AC V42934;  
DT 21-OCT-1998 (first entry)  
DE cDNA encoding murine neuroD2 protein, which is a bHLH protein.  
KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;  
KW classification; medulloblastoma; mouse; ss.  
OS Mus musculus.  
FT Key Location/Qualifiers  
FT cds 230..1378  
FT /\*tag= a  
FT /product= neuroD2  
PN US5795723-A.  
PD 18-AUG-1998.  
PF 07-AUG-1997; 910973.  
PR 07-AUG-1997; US-910973.  
PR 06-MAY-1994; US-239238.  
PR 02-NOV-1995; US-552142.  
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PI Olson JM, Tapscott SJ;  
DR P-PSDB; W71018.  
PT Classifying neuroectodermal tumours from expression pattern of



Result No.	Score	Query Match	Length	DB ID	Description
1	810	100.0	810	1 X05735	Human beta-amyloid
2	436.8	53.9	440	1 X41191	Human secreted pro
3	425.2	52.5	455	1 X41259	Human secreted pro
4	36.4	4.5	1951	1 T74893	Mouse neurogenic d
5	36.4	4.5	1951	1 T42934	CDNA encoding muri
6	35.4	4.4	110000	1 T58840_3	Continuation (4 of
7	34.6	4.3	4673	1 Q71189	P. yoelii SSP2 ant
8	33.4	4.1	1624	1 N91112	Polygalacturonase
9	33.4	4.1	1637	1 N80487	Plasmid clone pTcm
10	33.4	4.1	1636	1 Q38415	Polygalacturonase
11	33.4	4.1	9179	1 V74411	Staphylococcus aur
12	33	4.1	591	1 Q90286	EST clone BJ238. N
13	32.8	4.0	988	1 Q95079	Human flt-3 ligand
14	32.8	4.0	988	1 H81506	Human flt-3 ligand
15	32.8	4.0	3250	1 X00726	Human aggrecan deg
16	32.6	4.0	6524	1 T77803	CDNA encoding wild
17	32.6	4.0	917	1 X22226	Human secreted pro
18	32.4	4.0	8169	1 V28609	Actinomadura hibis
19	32.2	4.0	15672	1 Q10613	Rianodin receptor
20	32.2	4.0	3012	1 Q21188	Sequence encoding
21	32.2	4.0	715	1 X20370	Borrelia burgdorfe
22	32	4.0	110000	1 T21209_12	Continuation (13 o
23	32	4.0	13086	1 V74327	Staphylococcus aur
24	32	4.0	114955	1 X53491	Human adenosine Al
25	31.8	3.9	110000	1 X0248_01	Continuation (2 of
26	31.6	3.9	1075	1 Q64676	Human fibrillarlin
27	31.6	3.9	857	1 Q79462	Human S86/S109 flt
28	31.6	3.9	532	1 T67107	Helminth MIF CDNA
29	31.6	3.9	345	1 T67108	Helminth MIF CDNA
30	31.6	3.9	355	1 T67118	Helminth MIF PCR c
31	31.6	3.9	343	1 T67119	Helminth MIF CDNA
32	31.6	3.9	333	1 T67120	Helminth MIF PCR c
33	31.6	3.9	330	1 T67121	Helminth MIF CDNA
34	31.4	3.9	8700	1 Q42541	BgIII/HpaII fragme
35	31.4	3.9	8298	1 Q50168	p150 protein gene.
36	31.2	3.9	18079	1 N70753	Sequence of segmen
37	31.2	3.9	24379	1 T93095	Streptomyces freno
38	31.2	3.9	24379	1 T93095	Streptomyces roseo
39	31.2	3.9	1644	1 V69834	Human neural cell

Db 181 TCCAAGATGGCGCGCGCTGGCCGT

181 TCCAAGATGGCGCGCCTGGCCGTCTGGTCCGAGGGCCGTGACGGCCAGA 240



```
re-examination
JOURNAL Mol. Immunol. 28 (4-5), 449-457 (1991)
MEDLINE 91287731
COMMENT See also X61312-4 & X65266.
FEATURES Location/Qualifiers
1..1851
/organism="Carassius auratus"
/db_xref="taxon:7957"
/germline
/dev_stage="adult"
/clone_lib="GEM11 genomic"
/clone="6X"
358..365
misc_feature /note="promoter octamer"
633..722
misc_feature /note="FR1"
723..737
misc_feature /note="CDR1"
738..779
misc_feature /note="FR2"
780..836
misc_feature /note="CDR2"
837..935
misc_feature /note="FR3"
937..944
misc_feature /note="7-mer"
937..974
misc_feature /note="recombination signal sequence"
966..974
misc_feature /note="9-mer"
BASE COUNT 610 a 353 c 313 g 575 t
ORIGIN
```

```
Query Match 2.3%; Score 19; DB 4; Length 1851;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 AATGAACATTAGAAAAA 793
|||||
DB 1666 AATGAACATTAGAAAAA 1648
```

```
RESULT 15
AF194825/C
LOCUS AF194825 2069 bp DNA PLN 18-NOV-1999
DEFINITION Bougainvillea alba NADH dehydrogenase (ndhf) gene, partial cds;
chloroplast gene for chloroplast product.
ACCESSION AF194825
VERSION AF194825.1 GI:6449168
KEYWORDS
SOURCE
ORGANISM Bougainvillea alba.
Chloroplast Bougainvillea alba
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Caryophyllales; Nyctaginaceae;
Bougainvillea.
1 (bases 1 to 2069)
REFERENCE
AUTHORS Applequist, W.L. and Wallace, R.S.
TITLE Phylogeny of the portulacaceous cohort based on ndhf sequence data
JOURNAL Unpublished
REFERENCE
AUTHORS Applequist, W.L. and Wallace, R.S.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1999) Botany, Iowa State University, 353 Bessey
Hall, Ames, IA 50011-1020, USA
FEATURES Location/Qualifiers
1..2069
/organism="Bougainvillea alba"
/chloroplast
/db_xref="taxon:107614"
<1..>2069
/gene="ndhf"
```

```
CDS
<1..>2069
/gene="ndhf"
/codon_start=3
/product="NADH dehydrogenase"
/protein_id="AAF08857.1"
/db_xref="GI:6449215"
/translation="LIGAGLLFFPIATKNLRIRWAFSSILVLSIVMIFSMKLSIOQIN
SNFYQVYVSWTINNDFSEFGYLDIPLTISMSMLTITVAIVLISYDNMSHDRGYL
RFFAYMFFNTSMGLVITSSNLIQYIFWELVGMCSYLLIGFWFTRPFAANACQKAFV
TNRIGDFGLLGLVWITGTSFEGDLFEIFNNLNKNEINFLFCILCTLFLFAGPV
AKSAQFPLHWLPDAMEGPTPISALIHATWVAAGIFLVARLLPLFIVIPYTMNVLISF
IGIITVLLGATLALAQDKIKRSLAYSTMSQIGYMLALGMSIRTALEHLITHATYSKA
LLEFLGSGSIITHSMEAIVGSPDKQNMVLMGLARHVPITTKASFLLIGLSLGGIPPLA
CFWSKDEILNDSWLYSPIFAIAYTAGITAFYFRYIYLLTFFEGHLYLKNYSQKKS
NSFYISLWGERLPTINKNFCLLTLIMNNKRAFSTKTKKHQIDGNFKMKQSPF
ITMTYFYNNKVSYPHESENTMLPFLIILITFLTFLGIPVNOEGIDILTAKWLTSP
INLLHENLHNFLDWYEFVINAIVSVSIAYLGIFFLAFVLYKPKVYSSEKFNFLINSFDKI
GSKRIFWDKILYIYNWSSNGYIDSFYATSLIRGVRGLAELVIEFFDKRIIDGIPNGF
GVTSEFFV"
BASE COUNT 606 a 276 c 343 g 844 t
ORIGIN
```

Query Match 2.3%; Score 19; DB 8; Length 2069;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AATTGATTTCATCTATT 690  
|||||  
DB 609 AATTGATTTCATCTATT 591

Search completed: February 11, 2000, 05:40:18  
Job time: 904 sec



```
* * gap of unknown length
* * 13373: contig of 773 bp in length
* * gap of unknown length
* * 14175: contig of 802 bp in length
* * gap of unknown length
* * 14963: contig of 788 bp in length
* * gap of unknown length
* * 14964: contig of 798 bp in length
* * gap of unknown length
* * 15762: contig of 817 bp in length
* * gap of unknown length
* * 16579: contig of 811 bp in length
* * gap of unknown length
* * 17390: contig of 800 bp in length
* * gap of unknown length
* * 18190: contig of 786 bp in length
* * gap of unknown length
* * 18976: contig of 779 bp in length
* * gap of unknown length
* * 19755: contig of 759 bp in length
* * gap of unknown length
* * 20514: contig of 792 bp in length
* * gap of unknown length
* * 21306: contig of 794 bp in length
* * gap of unknown length
* * 22100: contig of 794 bp in length
* * gap of unknown length
* * 22894: contig of 787 bp in length
* * gap of unknown length
* * 23681: contig of 820 bp in length
* * gap of unknown length
* * 24501: contig of 773 bp in length
* * gap of unknown length
* * 25274: contig of 780 bp in length
* * gap of unknown length
* * 26054: contig of 755 bp in length
* * gap of unknown length
* * 26809: contig of 752 bp in length
* * gap of unknown length
* * 27561: contig of 819 bp in length
* * gap of unknown length
* * 28380: contig of 801 bp in length
* * gap of unknown length
* * 29181: contig of 796 bp in length
* * gap of unknown length
* * 29977: contig of 788 bp in length
* * gap of unknown length
* * 30765: contig of 803 bp in length
* * gap of unknown length
* * 31568: contig of 808 bp in length
* * gap of unknown length
* * 32376: contig of 778 bp in length
* * gap of unknown length
* * 33154: contig of 790 bp in length
* * gap of unknown length
* * 33944: contig of 772 bp in length
* * gap of unknown length
* * 34716: contig of 778 bp in length
* * gap of unknown length
* * 35494: contig of 791 bp in length
* * gap of unknown length
* * 36285: contig of 780 bp in length
* * gap of unknown length
* * 37065: contig of 811 bp in length
* * gap of unknown length
* * 37876: contig of 777 bp in length
* * gap of unknown length
* * 38653: contig of 768 bp in length
* * gap of unknown length
* * 39421: contig of 772 bp in length
* * gap of unknown length
* * 40193: contig of 781 bp in length
* * gap of unknown length
```

```
* * 40974 41754: contig of 781 bp in length
* * gap of unknown length
* * 41755 42558: contig of 804 bp in length
* * gap of unknown length
* * 42559 43360: contig of 802 bp in length
* * gap of unknown length
* * 43361 44170: contig of 810 bp in length
* * gap of unknown length
* * 44171 44965: contig of 795 bp in length
* * gap of unknown length
* * 44966 45756: contig of 791 bp in length
* * gap of unknown length
* * 45757 46543: contig of 787 bp in length
* * gap of unknown length
* * 46544 47348: contig of 805 bp in length
* * gap of unknown length
* * 47349 48152: contig of 804 bp in length
* * gap of unknown length
* * 48153 48951: contig of 799 bp in length
* * gap of unknown length
* * 48952 49743: contig of 792 bp in length
* * gap of unknown length
* * 49744 50547: contig of 804 bp in length
* * gap of unknown length
* * 50548 51324: contig of 777 bp in length
* * gap of unknown length
* * 51325 52121: contig of 797 bp in length
* * gap of unknown length
* * 52122 52894: contig of 773 bp in length
* * gap of unknown length
* * 52895 53677: contig of 783 bp in length
* * gap of unknown length
* * 53678 54458: contig of 781 bp in length
* * gap of unknown length
* * 54459 55225: contig of 767 bp in length
* * gap of unknown length
* * 55226 55972: contig of 747 bp in length
* * gap of unknown length
* * 55973 56789: contig of 817 bp in length
* * gap of unknown length
* * 56790 57605: contig of 816 bp in length
* * gap of unknown length
```

Query Match 2.5% Score 20; DB 44; Length 60013;  
Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGGCGCGCGCTGGCGGTC 206

Db 13672 ATGGCGCGCGCTGGCGGTC 13653

RESULT 14

CAVHH6X/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

CAVHH6X 1851 bp DNA VRT  
C.auratus immunoglobulin V-H hybridizing region (5X clone).  
X65266  
X65266.1 GI:62591  
immunoglobulin heavy chain variable region; V-H hybridizing region;  
goldfish.

Carassius auratus  
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprininae; Cyprinidae; Cyprinidae; Carassius.  
1 (bases 1 to 1851)  
Wilson, M.R.  
Direct Submission

Submitted (27-MAR-1992) M.R. Wilson, Basel Inst of Immunology,  
Grenzacherstrasse 487, Basel, SWITZERLAND  
2 (bases 1 to 1851)  
Wilson, M.R., Middleton, D. and Warr, G.W.  
Immunoglobulin VH genes of the goldfish, Carassius auratus: a

JOURNAL  
REFERENCE  
AUTHORS

*	39397	40222:	contig of 826 bp in length	gap of unknown length
*			gap of unknown length	
*	40223	41040:	contig of 818 bp in length	gap of unknown length
*			gap of unknown length	
*	41041	41854:	contig of 814 bp in length	gap of unknown length
*			gap of unknown length	
*	41855	42685:	contig of 831 bp in length	gap of unknown length
*			gap of unknown length	
*	42686	43496:	contig of 811 bp in length	gap of unknown length
*			gap of unknown length	
*	43497	44318:	contig of 822 bp in length	gap of unknown length
*			gap of unknown length	
*	44319	45133:	contig of 815 bp in length	gap of unknown length
*			gap of unknown length	
*	45134	45974:	contig of 841 bp in length	gap of unknown length
*			gap of unknown length	
*	45975	46759:	contig of 785 bp in length	gap of unknown length
*			gap of unknown length	
*	46760	47564:	contig of 805 bp in length	gap of unknown length
*			gap of unknown length	
*	47565	48379:	contig of 815 bp in length	gap of unknown length
*			gap of unknown length	
*	48380	49198:	contig of 819 bp in length	gap of unknown length
*			gap of unknown length	
*	49199	50012:	contig of 814 bp in length	gap of unknown length
*			gap of unknown length	
*	50013	50782:	contig of 770 bp in length	gap of unknown length
*			gap of unknown length	
*	50783	51575:	contig of 793 bp in length	gap of unknown length
*			gap of unknown length	
*	51576	52358:	contig of 783 bp in length	gap of unknown length
*			gap of unknown length	
*	52359	53161:	contig of 803 bp in length	gap of unknown length
*			gap of unknown length	
*	53162	53988:	contig of 827 bp in length	gap of unknown length
*			gap of unknown length	
*	53989	54797:	contig of 809 bp in length	gap of unknown length
*			gap of unknown length	
*	54798	55665:	contig of 868 bp in length	gap of unknown length
*			gap of unknown length	
*	55666	56479:	contig of 814 bp in length	gap of unknown length
*			gap of unknown length	
*	56480	57266:	contig of 787 bp in length	gap of unknown length
*			gap of unknown length	
*	57267	58073:	contig of 807 bp in length	gap of unknown length
*			gap of unknown length	
*	58074	58885:	contig of 812 bp in length	gap of unknown length
*			gap of unknown length	

```
Query Match      2.5%; Score 20; DB 44; Length 67795;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT	l3
LOCUS	AC016814/c
DEFINITION	AC016814 60013 bp DNA HTG 07-DEC-1999
ACCESSION	Mus musculus clone RP23-2M16, LOW-PASS SEQUENCE SAMPLING.
VERSION	AC016814
KEYWORDS	AC016814.1 GI:6532079
SOURCE	HTG; HTGS_PHASED.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Craniata; Vertebrata; Mammalia;
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 60013)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Mus musculus, clone RP23-2M16

Unpublished  
2 (bases 1

Barren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Dearluzh, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehotsky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., McDermid, J., Morrow, J., Taylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassilev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1965, Vol. 68, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1965, Vol. 68, No. 3, pp. 21-30
4. The Effect of Student Motivation on Learning	Journal of Educational Research	1965, Vol. 68, No. 4, pp. 31-40
5. The Role of the School in Society	Journal of Educational Research	1965, Vol. 68, No. 5, pp. 41-50
6. The Impact of Teacher Education on Classroom Practice	Journal of Educational Research	1965, Vol. 68, No. 6, pp. 51-60
7. The Importance of Professional Development	Journal of Educational Research	1965, Vol. 68, No. 7, pp. 61-70
8. The Effect of Curriculum Reform on Student Achievement	Journal of Educational Research	1965, Vol. 68, No. 8, pp. 71-80
9. The Role of the School in Promoting Social Change	Journal of Educational Research	1965, Vol. 68, No. 9, pp. 81-90
10. The Impact of Teacher Attitudes on Student Behavior	Journal of Educational Research	1965, Vol. 68, No. 10, pp. 91-100

```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 67795)
JOURNAL     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE   Homo sapiens, clone RP11-10G1
AUTHORS     2 (bases 1 to 67795)
TITLE       Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
AUTHORS     Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
AUTHORS     Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
AUTHORS     Cooke, P., DeAvellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
AUTHORS     Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
AUTHORS     Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
AUTHORS     Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
AUTHORS     Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
AUTHORS     McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
AUTHORS     Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
AUTHORS     Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
AUTHORS     Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
AUTHORS     Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
AUTHORS     Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
COMMENT     Direct Submission
TITLE       Submitted (03-DEC-1999) Whitehead Institute/MIT Center for Genome
COMMENT     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
COMMENT     Smit, A.F.A. & Green, P. (1996-1997)
COMMENT     http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT     ----- Genome Center
TITLE       Center: Whitehead Institute/ MIT Center for Genome Research
COMMENT     Center code: WIBR
COMMENT     Web site: http://www-seq.wi.mit.edu
COMMENT     Contact: sequence_submission@genome.wi.mit.edu
COMMENT     ----- Project Information
TITLE       Center project name: L3059
COMMENT     Center clone name: 10_G_1
COMMENT     -----
* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 790: contig of 790 bp in length
*      gap of unknown length
* 791 1599: contig of 799 bp in length
*      gap of unknown length
* 1590 2411: contig of 822 bp in length
*      gap of unknown length
* 2412 3206: contig of 795 bp in length
*      gap of unknown length
* 3207 4010: contig of 804 bp in length
*      gap of unknown length
* 4011 4799: contig of 789 bp in length
*      gap of unknown length
* 4800 5607: contig of 808 bp in length
*      gap of unknown length
* 5608 6417: contig of 810 bp in length
*      gap of unknown length
* 6418 7207: contig of 790 bp in length
*      gap of unknown length
* 7208 8023: contig of 816 bp in length
*      gap of unknown length
* 8024 8833: contig of 810 bp in length
*      gap of unknown length
* 8834 9641: contig of 808 bp in length
*      gap of unknown length
*
* 10443: contig of 802 bp in length
*      gap of unknown length
* 10444 11240: contig of 797 bp in length
*      gap of unknown length
* 11241 11974: contig of 734 bp in length
*      gap of unknown length
* 11975 12772: contig of 798 bp in length
*      gap of unknown length
* 12773 13577: contig of 805 bp in length
*      gap of unknown length
* 13578 14393: contig of 816 bp in length
*      gap of unknown length
* 14394 15136: contig of 743 bp in length
*      gap of unknown length
* 15137 15949: contig of 813 bp in length
*      gap of unknown length
* 15950 16770: contig of 821 bp in length
*      gap of unknown length
* 16771 17577: contig of 807 bp in length
*      gap of unknown length
* 17578 18351: contig of 774 bp in length
*      gap of unknown length
* 18352 19153: contig of 802 bp in length
*      gap of unknown length
* 19154 19956: contig of 803 bp in length
*      gap of unknown length
* 19957 20766: contig of 810 bp in length
*      gap of unknown length
* 20767 21588: contig of 822 bp in length
*      gap of unknown length
* 21589 22405: contig of 817 bp in length
*      gap of unknown length
* 22406 23225: contig of 820 bp in length
*      gap of unknown length
* 23226 24052: contig of 827 bp in length
*      gap of unknown length
* 24053 24873: contig of 821 bp in length
*      gap of unknown length
* 24874 25686: contig of 813 bp in length
*      gap of unknown length
* 25687 26497: contig of 811 bp in length
*      gap of unknown length
* 26498 27313: contig of 816 bp in length
*      gap of unknown length
* 27314 28113: contig of 800 bp in length
*      gap of unknown length
* 28114 28895: contig of 782 bp in length
*      gap of unknown length
* 28896 29709: contig of 814 bp in length
*      gap of unknown length
* 29710 30526: contig of 817 bp in length
*      gap of unknown length
* 30527 31339: contig of 813 bp in length
*      gap of unknown length
* 31340 32149: contig of 810 bp in length
*      gap of unknown length
* 32150 32955: contig of 806 bp in length
*      gap of unknown length
* 32956 33770: contig of 815 bp in length
*      gap of unknown length
* 33771 34582: contig of 812 bp in length
*      gap of unknown length
* 34583 35406: contig of 824 bp in length
*      gap of unknown length
* 35407 36226: contig of 820 bp in length
*      gap of unknown length
* 36227 37023: contig of 797 bp in length
*      gap of unknown length
* 37024 37804: contig of 781 bp in length
*      gap of unknown length
* 37805 38579: contig of 775 bp in length
*      gap of unknown length
* 38580 39396: contig of 817 bp in length
*      gap of unknown length

```

```
* * gap of unknown length
* * 7756: contig of 774 bp in length
* * gap of unknown length
* * 8556: contig of 800 bp in length
* * gap of unknown length
* * 9339: contig of 783 bp in length
* * gap of unknown length
* * 10124: contig of 785 bp in length
* * gap of unknown length
* * 10901: contig of 777 bp in length
* * gap of unknown length
* * 11678: contig of 777 bp in length
* * gap of unknown length
* * 12436: contig of 758 bp in length
* * gap of unknown length
* * 13215: contig of 779 bp in length
* * gap of unknown length
* * 13996: contig of 781 bp in length
* * gap of unknown length
* * 14785: contig of 789 bp in length
* * gap of unknown length
* * 15573: contig of 788 bp in length
* * gap of unknown length
* * 16356: contig of 783 bp in length
* * gap of unknown length
* * 17133: contig of 777 bp in length
* * gap of unknown length
* * 17913: contig of 780 bp in length
* * gap of unknown length
* * 18697: contig of 784 bp in length
* * gap of unknown length
* * 19486: contig of 789 bp in length
* * gap of unknown length
* * 20258: contig of 772 bp in length
* * gap of unknown length
* * 21029: contig of 771 bp in length
* * gap of unknown length
* * 21806: contig of 777 bp in length
* * gap of unknown length
* * 22578: contig of 772 bp in length
* * gap of unknown length
* * 23351: contig of 773 bp in length
* * gap of unknown length
* * 24109: contig of 758 bp in length
* * gap of unknown length
* * 24887: contig of 778 bp in length
* * gap of unknown length
* * 25666: contig of 779 bp in length
* * gap of unknown length
* * 26440: contig of 774 bp in length
* * gap of unknown length
* * 27215: contig of 775 bp in length
* * gap of unknown length
* * 27995: contig of 780 bp in length
* * gap of unknown length
* * 28790: contig of 795 bp in length
* * gap of unknown length
* * 29560: contig of 770 bp in length
* * gap of unknown length
* * 30341: contig of 781 bp in length
* * gap of unknown length
* * 31111: contig of 770 bp in length
* * gap of unknown length
* * 31892: contig of 781 bp in length
* * gap of unknown length
* * 32673: contig of 781 bp in length
* * gap of unknown length
* * 33452: contig of 779 bp in length
* * gap of unknown length
* * 34237: contig of 785 bp in length
* * gap of unknown length
* * 34997: contig of 760 bp in length
* * gap of unknown length
```

```
* * 34998 35774: contig of 777 bp in length
* * gap of unknown length
* * 35775 36560: contig of 786 bp in length
* * gap of unknown length
* * 36561 37341: contig of 781 bp in length
* * gap of unknown length
* * 37342 38101: contig of 760 bp in length
* * gap of unknown length
* * 38102 38895: contig of 794 bp in length
* * gap of unknown length
* * 38896 39674: contig of 779 bp in length
* * gap of unknown length
* * 39675 40406: contig of 732 bp in length
* * gap of unknown length
* * 40407 41186: contig of 780 bp in length
* * gap of unknown length
* * 41187 41966: contig of 780 bp in length
* * gap of unknown length
* * 41967 42696: contig of 730 bp in length
* * gap of unknown length
* * 42697 43458: contig of 762 bp in length
* * gap of unknown length
* * 43459 44247: contig of 789 bp in length
* * gap of unknown length
* * 44248 45027: contig of 780 bp in length
* * gap of unknown length
* * 45028 45813: contig of 786 bp in length
* * gap of unknown length
* * 45814 46610: contig of 797 bp in length
* * gap of unknown length
* * 46611 47377: contig of 767 bp in length
* * gap of unknown length
* * 47378 48138: contig of 761 bp in length
* * gap of unknown length
* * 48139 48920: contig of 782 bp in length
* * gap of unknown length
* * 48921 49698: contig of 778 bp in length
* * gap of unknown length
* * 49699 50482: contig of 784 bp in length
* * gap of unknown length
* * 50483 51269: contig of 787 bp in length
* * gap of unknown length
* * 51270 52043: contig of 774 bp in length
* * gap of unknown length
* * 52044 52797: contig of 754 bp in length
* * gap of unknown length
* * 52798 53596: contig of 799 bp in length
* * gap of unknown length
* * 53597 54394: contig of 798 bp in length
* * gap of unknown length
* * 54395 55172: contig of 778 bp in length
* * gap of unknown length
* * 55173 55953: contig of 781 bp in length
* * gap of unknown length
* * 55954 56730: contig of 777 bp in length
* * gap of unknown length
```

Query Match 2.5%; Score 20; DB 44; Length 62216;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGGCGCGCGCTGGCCGTC 206

Db 12430 ATGGCGCGCGCTGGCCGTC 12411

RESULT 12

AC016538  
LOCUS AC016538 67795 bp DNA HTG 03-DEC-1999  
DEFINITION Homo sapiens clone RP11-10G1, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC016538  
VERSION AC016538.1 GI:6513994  
KEYWORDS HTG; HTGS\_PHASE0.

```

repeat_region /note="5 copies of 6 mer 93 % conserved"
55255..55531
/note="AluSq repeat: matches 1. .278 of consensus;
incomplete repeat"
repeat_region 55532..55567
/note="18 copies of 2 mer 86 % conserved"
56142..56233
/note="MIR repeat: matches 52. .141 of consensus"
57325..57491
/note="AluJo repeat: matches 136. .302 of consensus;
incomplete repeat"
repeat_region 58111..58792
/note="AluJo repeat: matches 300. .13 of consensus"
59799..59965
/note="MIR repeat: matches 213. .47 of consensus"
60865..60900
/note="9 copies of 4 mer 83 % conserved"
61015..61093
/note="MER6 repeat: matches 866. .785 of consensus"
61082..61156
/note="MER6 repeat: matches 75. .1 of consensus"
61153..61223
/note="MER6 repeat: matches 797. .866 of consensus"
61229..61311
/note="MIR repeat: matches 198. .104 of consensus"
63776..64004
/note="MIR repeat: matches 262. .34 of consensus"
64717..64911
/note="MIR repeat: matches 257. .68 of consensus"
65581..65898
/note="LIMC3 repeat: matches 2156. .2483 of consensus"
65834..65919
/note="LIMC1 repeat: matches 998. .1079 of consensus"
67174..67428
/note="MIR repeat: matches 5. .262 of consensus"
67616..67860
/note="MIR repeat: matches 262. .2 of consensus"
67865..68027
/note="MIR repeat: matches 262. .82 of consensus"
70575..70610
/note="9 copies of 4 mer 92 % conserved"
71068..71230
/note="LIPAS repeat: matches 730. .892 of consensus"
72756..73649
/note="LIME2 repeat: matches 10. .911 of consensus"
73770..73820
/note="MER5A repeat: matches 167. .117 of consensus"
74476..74643
/note="MIR repeat: matches 88. .262 of consensus"
74962..75050
/note="MIR2 repeat: matches 60. .145 of consensus"
77600..77633
/note="17 copies of 2 mer 85 % conserved"
81606..81779
/note="MER45 repeat: matches 177. .1 of consensus"
complement(<82028..85660)
/note="match: D16473 mRNA, Xq terminal portion"
83381..83406
/note="13 copies of 2 mer 100 % conserved"
83939..84598
/note="LIMC1 repeat: matches 394. .1079 of consensus"
84641..84764
/note="MIR repeat: matches 188. .74 of consensus"

Query Match 2.5%; Score 20; DB 11; Length 98209;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 108 CTTGCGCCCTGTTGCCCTTCT 127

|||||

DB 52274 CTTGCGCCCTGTTGCCCTTCT 52293

RESULT 11  
AC016414/c  
LOCUS  
DEFINITION  
AC016414  
VERSION  
AC016414.1 GI:6468838  
KEYWORDS  
HTG; HTGS\_PHASE0.  
SOURCE  
human.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AC016414 62216 bp DNA HTG 26-NOV-1999  
Homo sapiens clone RP11-10A17, LOW-PASS SEQUENCE SAMPLING.  
AC016414

AC016414.1 GI:6468838

HTG; HTGS\_PHASE0.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 62216)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-10A17

2 (bases 1 to 62216)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,

Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Saverly,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (26-NOV-1999), Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit,A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submission@genome.wi.mit.edu

----- Project Information

Center project name: L3030

Center clone name: 10\_A\_17

-----

\* NOTE: This record contains 80 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 742: contig of 742 bp in length

\* 743 gap of unknown length

\* 1523: contig of 781 bp in length

\* gap of unknown length

\* 1524 contig of 777 bp in length

\* gap of unknown length

\* 2301 contig of 775 bp in length

\* gap of unknown length

\* 3076 contig of 767 bp in length

\* gap of unknown length

\* 3843 contig of 773 bp in length

\* gap of unknown length

\* 4616 contig of 769 bp in length

\* gap of unknown length

\* 5385 contig of 781 bp in length

\* gap of unknown length

\* 6166 contig of 817 bp in length



```

* 97731 97810: gap of unknown length
* 97811 98334: contig of 524 bp in length
* 98335 98414: gap of unknown length
* 98415 98985: contig of 571 bp in length
* 98986 99065: gap of unknown length
* 99066 99291: contig of 226 bp in length
* 99292 99371: gap of unknown length
* 99372 99912: contig of 541 bp in length
* 99913 99992: gap of unknown length
* 99993 100618: contig of 626 bp in length
* 100619 100698: gap of unknown length
* 100699 101398: contig of 700 bp in length

Query Match 2.6%; Score 21; DB 42; Length 113562;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 AGCGGAAGTGGCGGAGAAAGTG 174
|||||
1865 AGCGGAAGTGGCGGAGAAAGTG 1885

RESULT 7
AC012995/c
LOCUS AC012995 76396 bp DNA HTG 03-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC012995
VERSION AC012995.1 GI:6223070
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 76396)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212198 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..76396
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 21300 a 16411 c 16836 g 21849 t
ORIGIN

Query Match 2.6%; Score 21; DB 43; Length 76396;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 CTGGCGGAGCGGAGCGG 158
|||||
Db 12113 CTGGCGGAGCGGAGCGG 12093

RESULT 8
AC017975
LOCUS AC017975 27387 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017975
VERSION AC017975.1 GI:6553215
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.

```

```

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 27387)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212858 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..27387
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 7801 a 5791 c 5937 g 7858 t
ORIGIN

Query Match 2.6%; Score 21; DB 45; Length 27387;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 AGCGGAAGTGGCGGAGAAAGTG 174
|||||
Db 19673 AGCGGAAGTGGCGGAGAAAGTG 19693

RESULT 9
AP000076
LOCUS AP000076 100000 bp DNA PRI 20-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 8p11.2, senescence gene
region, section 12/19, complete sequence.
ACCESSION AP000076
VERSION AP000076.1 GI:4579897
KEYWORDS HTG.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Isumura, M., Ikegawa, S., Kinjo, T. and Nakamura, Y.
DNA sequence analysis of a 1.9-Mb region on chromosome 8p11.2
Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 100000)
Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, J.
Direct Submission
Submitted (12-FEB-1999) to the DDBJ/EMBL/GenBank databases. Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Database Department; 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
>>text_table: 0000140177.
FEATURES
source
1..100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p11.2"
43625..43766
/note="SHGC-9028; The location is between each flanking
site of PCR primers."
/db_xref="GDB:1235120"

BASE COUNT 31453 a 18346 c 19549 g 30651 t
ORIGIN

Query Match 2.5%; Score 20; DB 10; Length 100000;
Best Local Similarity 100.0%; Pred. No. 2.2;

```

the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 93 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 964: contig of 964 bp in length  
\* 965 1044: gap of unknown length  
\* 1045 2062: contig of 1018 bp in length  
\* 2063 2142: gap of unknown length  
\* 2143 2986: contig of 844 bp in length  
\* 2987 3066: gap of unknown length  
\* 3067 3634: contig of 568 bp in length  
\* 3635 3714: gap of unknown length  
\* 3715 4639: contig of 924 bp in length  
\* 4639 5291: gap of unknown length  
\* 5291 5371: contig of 573 bp in length  
\* 5372 6106: contig of 735 bp in length  
\* 6107 6186: gap of unknown length  
\* 6187 7016: contig of 830 bp in length  
\* 7017 7096: gap of unknown length  
\* 7097 8094: contig of 998 bp in length  
\* 8095 8174: gap of unknown length  
\* 8175 9007: contig of 833 bp in length  
\* 9008 9087: gap of unknown length  
\* 9088 10029: contig of 942 bp in length  
\* 10030 10109: gap of unknown length  
\* 10110 11068: contig of 959 bp in length  
\* 11069 11148: gap of unknown length  
\* 11149 11725: contig of 577 bp in length  
\* 11726 11805: gap of unknown length  
\* 11806 12392: contig of 590 bp in length  
\* 12396 12475: gap of unknown length  
\* 12476 13403: contig of 928 bp in length  
\* 13404 13483: gap of unknown length  
\* 13484 14236: contig of 753 bp in length  
\* 14237 14316: gap of unknown length  
\* 14317 15330: contig of 1014 bp in length  
\* 15331 15410: gap of unknown length  
\* 15411 16745: contig of 1335 bp in length  
\* 16746 16825: gap of unknown length  
\* 16826 17681: contig of 856 bp in length  
\* 17682 17761: gap of unknown length  
\* 17762 18569: contig of 808 bp in length  
\* 18570 18649: gap of unknown length  
\* 18650 19700: contig of 1051 bp in length  
\* 19701 19780: gap of unknown length  
\* 19781 21144: contig of 1364 bp in length  
\* 21145 21224: gap of unknown length  
\* 21225 21905: contig of 681 bp in length  
\* 21906 21985: gap of unknown length  
\* 21986 22953: contig of 968 bp in length  
\* 22954 23033: gap of unknown length  
\* 23034 24086: contig of 1053 bp in length  
\* 24087 24166: gap of unknown length  
\* 24167 25883: contig of 1717 bp in length  
\* 25884 25963: gap of unknown length  
\* 25964 27209: contig of 1246 bp in length  
\* 27210 27289: gap of unknown length  
\* 27290 28436: contig of 1147 bp in length  
\* 28437 28516: gap of unknown length  
\* 28517 29905: contig of 1389 bp in length  
\* 29906 29985: gap of unknown length  
\* 29986 31609: contig of 1624 bp in length  
\* 31610 31689: gap of unknown length  
\* 31690 32620: contig of 931 bp in length  
\* 32621 32700: gap of unknown length  
\* 32701 33457: contig of 757 bp in length  
\* 33458 33537: gap of unknown length

33538 34762: contig of 1225 bp in length  
\* 34763 34842: gap of unknown length  
\* 34843 35650: contig of 808 bp in length  
\* 35651 3730: gap of unknown length  
\* 3731 37654: contig of 1924 bp in length  
\* 37655 37734: gap of unknown length  
\* 37735 38356: contig of 622 bp in length  
\* 38357 38436: gap of unknown length  
\* 38437 40525: contig of 2089 bp in length  
\* 40526 40605: gap of unknown length  
\* 40606 42579: contig of 1374 bp in length  
\* 42580 42659: gap of unknown length  
\* 42660 44115: contig of 1456 bp in length  
\* 44116 44195: gap of unknown length  
\* 44196 45369: contig of 1174 bp in length  
\* 45370 45449: gap of unknown length  
\* 45450 47205: contig of 1756 bp in length  
\* 47206 47285: gap of unknown length  
\* 47286 49223: contig of 1938 bp in length  
\* 49224 49303: gap of unknown length  
\* 49304 51498: contig of 2195 bp in length  
\* 51499 51578: gap of unknown length  
\* 51579 52672: contig of 1094 bp in length  
\* 52673 52752: gap of unknown length  
\* 52753 54833: contig of 2081 bp in length  
\* 54834 54913: gap of unknown length  
\* 54914 56621: contig of 1708 bp in length  
\* 56622 56701: gap of unknown length  
\* 56702 58223: contig of 1522 bp in length  
\* 58224 58303: gap of unknown length  
\* 58304 59080: contig of 777 bp in length  
\* 59081 59160: gap of unknown length  
\* 59161 60156: contig of 996 bp in length  
\* 60157 60236: gap of unknown length  
\* 60237 62470: contig of 2234 bp in length  
\* 62471 62550: gap of unknown length  
\* 62551 64581: contig of 2031 bp in length  
\* 64582 64661: gap of unknown length  
\* 64662 66722: contig of 2061 bp in length  
\* 66723 68802: gap of unknown length  
\* 68803 69042: contig of 2240 bp in length  
\* 69043 69122: gap of unknown length  
\* 69123 71027: contig of 1905 bp in length  
\* 71028 71107: gap of unknown length  
\* 71108 73332: contig of 2225 bp in length  
\* 73333 73412: gap of unknown length  
\* 73413 75475: contig of 2063 bp in length  
\* 75476 75555: gap of unknown length  
\* 75556 78424: contig of 2869 bp in length  
\* 78425 78504: gap of unknown length  
\* 78505 85099: contig of 6595 bp in length  
\* 85100 85179: gap of unknown length  
\* 85180 90487: contig of 5308 bp in length  
\* 90488 90567: gap of unknown length  
\* 90568 91182: contig of 615 bp in length  
\* 91183 91262: gap of unknown length  
\* 91263 91894: contig of 632 bp in length  
\* 91895 91974: gap of unknown length  
\* 91975 92461: contig of 487 bp in length  
\* 92462 92541: gap of unknown length  
\* 92542 93207: contig of 666 bp in length  
\* 93208 93287: gap of unknown length  
\* 93288 93741: contig of 454 bp in length  
\* 93742 93821: gap of unknown length  
\* 93822 95002: contig of 1181 bp in length  
\* 95003 95082: gap of unknown length  
\* 95083 95833: contig of 751 bp in length  
\* 95834 95913: gap of unknown length  
\* 95914 96439: contig of 526 bp in length  
\* 96440 96519: gap of unknown length  
\* 96520 97094: contig of 575 bp in length  
\* 97095 97174: gap of unknown length  
\* 97175 97730: contig of 556 bp in length



```

RESULT 5
DMU35074 3937 bp mRNA INV 17-APR-1998
LOCUS Drosophila melanogaster centrosomal and chromosomal factor (ccf)
DEFINITION mRNA, complete cds and 5' genomic segment.
ACCESSION U35074
VERSION U35074.1 GI:1016261
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 3937)
AUTHORS Kodjabachian, L., Delaage, M., Maurel, C., Miassod, R., Jacq, B. and
Rosset, R.
TITLE Mutations in ccf, a novel Drosophila gene encoding a chromosomal
factor, affect progression through mitosis and interact with Pc-G
mutations
JOURNAL EMBO J. 17 (4), 1063-1075 (1998)
EDLINE 98130597
REFERENCE 2 (bases 1 to 3937)
AUTHORS Kodjabachian, L.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-1995) Laurent Kodjabachian, Laboratoire de
Genetique et Physiologie du Developpement, CNRS, Parc Scientifique
de Luminy, Case 907, Marseille 13288, France
FEATURES
source
1..3937
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/map="3R; 82E"
1..425
/gene="corto"
/note="genomic DNA"
/db_xref="FlyBase:FBgn0010313"
1..3937
/gene="corto"
/note="corto"
/allele="n"
/db_xref="FlyBase:FBgn0010313"
293..298
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
407
/gene="corto"
/note="at the indicated position in the A46 enhancer trap
line is inserted a P(lacZ, ry) transposable element."
/db_xref="FlyBase:FBgn0010313"
426..1478
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
426..3937
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
585..590
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
635..640
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
1479..3131
/gene="corto"
/note="required for mitotic chromosome condensation;
Centrosomal and Chromosomal Factor"
/codon_start=1
/product="CCF"
/protein_id="AAC13917.1"
/db_xref="GI:1016262"
/translation="MTMAACVANYDMSSLHGMSSALSALQQQQQQQQQQSQTQQQHH
HQQQQHHYAAVAHQQLQQQQQHHRRHHQPANTSSSRHSHAAATQVA
AAVANSRQQQQQQQQQQQQQQQQQQTASSNSNTAPAPSPQKDYSLPHVDCSVYELPN

```

```

OPKPPAGQORVEPLLTHPCYFRKMESSORSPFVNNMHATRAVSSSSLSGAAAGGGG
SGAVATAPSSSARARGARATSAQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
QQQSSLVRRHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
AAGWQQLAALAAATALTTPHMLHPHSHYAAKSGSGGQGGKRDAMTSSGSGQTAVAS
KQQQSQVQQQQQQQQQQHCLPPPPMDATSMLEKDPATVPSNYQAGPDPNPNRLYS
ATPTAGAAAGSGSASVGGGAVSGAGATGAVNTATDKLSGKYRQYRLRSQRMHPYAAAS
LNLAAGAAAGQTSFVPFSSAATAVAATPTFQHLPLQISCYNV"
3'UTR 3132..3937
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
3811..3816
/gene="corto"
/db_xref="FlyBase:FBgn0010313"
3868
BASE COUNT 1293 a 1098 c 926 g 620 t
ORIGIN
Query Match 2.6%; Score 21; DB 34; Length 3937;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 CTTGGCGGCGGAGCGGAGCGG 158
|||||
DB 2117 CTTGGCGGCGGAGCGGAGCGG 2137
|||||
RESULT 6
AC011072 113562 bp DNA HTG 20-OCT-1999
LOCUS Drosophila melanogaster chromosome 2 clone BACR48A20 (D937) RPCI-98
DEFINITION 48.A.20 map 55B-55D strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
*** 93 unordered pieces.
AC011072
ACCESSION AC011072.4 GI:6087903
VERSION HTG; HTGS_PHASE1.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 113562)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 113562)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory MS 64-121, Berkeley, CA 94720, USA
COMMENT On Oct 20, 1999 this sequence version replaced gi:6067123.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu. All contigs in this submission meet

```

```
* 5970 6049: gap of unknown length
* 6050 6771: contig of 722 bp in length
* 6772 6851: gap of unknown length
* 6852 8136: contig of 1285 bp in length
* 8137 8216: gap of unknown length
* 8217 8879: contig of 663 bp in length
* 8880 8959: gap of unknown length
* 8960 9778: contig of 819 bp in length
* 9779 9858: gap of unknown length
* 9859 10614: contig of 756 bp in length
* 10615 11280: contig of 586 bp in length
* 11281 11360: gap of unknown length
* 11361 12029: contig of 669 bp in length
* 12030 12109: gap of unknown length
* 12110 12854: contig of 745 bp in length
* 12855 12934: gap of unknown length
* 12935 14181: contig of 1247 bp in length
* 14182 14261: gap of unknown length
* 14262 14972: contig of 711 bp in length
* 14973 15052: gap of unknown length
* 15053 15978: contig of 926 bp in length
* 15979 16058: gap of unknown length
* 16059 16721: contig of 663 bp in length
* 16722 16801: gap of unknown length
* 16802 17774: contig of 973 bp in length
* 17775 17854: gap of unknown length
* 17855 18669: contig of 815 bp in length
* 18670 18749: gap of unknown length
* 18750 19787: contig of 1038 bp in length
* 19788 19867: gap of unknown length
* 19868 21014: contig of 1147 bp in length
* 21015 21094: gap of unknown length
* 21095 22335: contig of 1141 bp in length
* 22336 22315: gap of unknown length
* 22316 23360: contig of 1045 bp in length
* 23361 23440: gap of unknown length
* 23441 24064: contig of 624 bp in length
* 24065 24144: gap of unknown length
* 24145 25411: contig of 1267 bp in length
* 25412 25491: gap of unknown length
* 25492 28415: contig of 924 bp in length
* 28416 28495: gap of unknown length
* 28496 27203: contig of 708 bp in length
* 27204 27283: gap of unknown length
* 27284 27908: contig of 625 bp in length
* 27909 27988: gap of unknown length
* 27989 29152: contig of 1164 bp in length
* 29153 29232: gap of unknown length
* 29233 29730: contig of 498 bp in length
* 29731 29810: gap of unknown length
* 29811 31005: contig of 1195 bp in length
* 31006 31085: gap of unknown length
* 31086 32359: contig of 1274 bp in length
* 32360 32439: gap of unknown length
* 32440 33158: contig of 719 bp in length
* 33159 33238: gap of unknown length
* 33239 33901: contig of 663 bp in length
* 33902 33981: gap of unknown length
* 33982 34630: contig of 649 bp in length
* 34631 34710: gap of unknown length
* 34711 35484: contig of 774 bp in length
* 35485 35564: gap of unknown length
* 35565 36639: contig of 1075 bp in length
* 36640 36719: gap of unknown length
* 36720 37755: contig of 1056 bp in length
* 37756 37856: gap of unknown length
* 37857 38578: contig of 723 bp in length
* 38579 38658: gap of unknown length
* 38659 39371: contig of 713 bp in length
* 39372 39451: gap of unknown length
* 39452 40368: contig of 917 bp in length
* 40369 40448: gap of unknown length
```

```
* 40449 41736: contig of 1288 bp in length
* 41737 41816: gap of unknown length
* 41817 43163: contig of 1347 bp in length
* 43164 43243: gap of unknown length
* 43244 44365: contig of 1122 bp in length
* 44366 44445: gap of unknown length
* 44446 45300: contig of 855 bp in length
* 45301 45380: gap of unknown length
* 45381 46514: contig of 1134 bp in length
* 46515 46594: gap of unknown length
* 46595 48237: contig of 1643 bp in length
* 48238 48317: gap of unknown length
* 48318 49655: contig of 1338 bp in length
* 49656 49735: gap of unknown length
* 49736 51205: contig of 1470 bp in length
* 51206 51285: gap of unknown length
* 51286 52492: contig of 1207 bp in length
* 52493 52572: gap of unknown length
* 52573 54279: contig of 1707 bp in length
* 54280 54359: gap of unknown length
* 54360 55688: contig of 1329 bp in length
* 55689 55768: gap of unknown length
* 55769 56577: contig of 889 bp in length
* 56578 56737: gap of unknown length
* 56738 58423: contig of 1686 bp in length
* 58424 58503: gap of unknown length
* 58504 60089: contig of 1586 bp in length
* 60090 60169: gap of unknown length
* 60170 61484: contig of 1315 bp in length
* 61485 61564: gap of unknown length
* 61565 62509: gap of unknown length
* 62510 63286: contig of 777 bp in length
* 63287 63366: gap of unknown length
* 63367 64268: contig of 902 bp in length
* 64269 64348: gap of unknown length
* 64349 66125: contig of 1777 bp in length
* 66126 66205: gap of unknown length
* 66206 68260: contig of 2055 bp in length
* 68261 68340: gap of unknown length
* 68341 70058: contig of 1718 bp in length
* 70059 70138: gap of unknown length
* 70139 71508: contig of 1370 bp in length
* 71509 71588: gap of unknown length
* 71589 73765: contig of 2177 bp in length
* 73766 73845: gap of unknown length
* 73846 75975: contig of 2130 bp in length
* 75976 76055: gap of unknown length
* 76056 77978: contig of 1923 bp in length
* 77979 78058: gap of unknown length
* 78059 80306: contig of 2248 bp in length
* 80307 80386: gap of unknown length
* 80387 84726: contig of 4339 bp in length
* 84726 84805: gap of unknown length
* 84806 88580: contig of 3775 bp in length
* 88581 88660: gap of unknown length
* 88661 94422: contig of 5762 bp in length
* 94423 94502: gap of unknown length
* 94503 95203: contig of 701 bp in length
* 95204 95283: gap of unknown length
* 95284 95887: contig of 604 bp in length
* 95888 95967: gap of unknown length
* 95968 96613: contig of 646 bp in length
```

Query Match 2.6%; Score 21; DB 33; Length 158018;  
Best Local Similarity 100.0%; Pred. No. 0.57;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ACCGGAAGTGGCGAGAAAGTG 174

|||||

DB 153844 ACCGGAAGTGGCGAGAAAGTG 153864

51509 51588: gap of unknown length  
 51589 53474: contig of 1886 bp in length  
 53475 53474: gap of unknown length  
 53476 53472: contig of 1918 bp in length  
 53477 53472: gap of unknown length  
 53478 53472: contig of 2057 bp in length  
 53479 53472: gap of unknown length  
 53480 53472: contig of 2620 bp in length  
 53481 53472: gap of unknown length  
 53482 53472: contig of 2496 bp in length  
 53483 53472: gap of unknown length  
 53484 53472: contig of 2704 bp in length  
 53485 53472: gap of unknown length  
 53486 53472: contig of 1757 bp in length  
 53487 53472: gap of unknown length  
 53488 53472: contig of 3603 bp in length  
 53489 53472: gap of unknown length  
 53490 53472: contig of 2803 bp in length  
 53491 53472: gap of unknown length  
 53492 53472: contig of 5394 bp in length  
 53493 53472: gap of unknown length  
 53494 53472: contig of 3629 bp in length  
 53495 53472: gap of unknown length  
 53496 53472: contig of 4039 bp in length  
 53497 53472: gap of unknown length  
 53498 53472: contig of 6171 bp in length  
 53499 53472: gap of unknown length  
 53500 53472: contig of 7622 bp in length  
 53501 53472: gap of unknown length  
 53502 53472: contig of 732 bp in length  
 53503 53472: gap of unknown length  
 53504 53472: contig of 642 bp in length  
 53505 53472: gap of unknown length  
 53506 53472: contig of 653 bp in length  
 53507 53472: gap of unknown length  
 53508 53472: contig of 705 bp in length  
 53509 53472: gap of unknown length  
 53510 53472: contig of 565 bp in length  
 53511 53472: gap of unknown length  
 53512 53472: contig of 507 bp in length  
 53513 53472: gap of unknown length  
 53514 53472: contig of 679 bp in length  
 53515 53472: gap of unknown length  
 53516 53472: contig of 584 bp in length  
 53517 53472: gap of unknown length  
 53518 53472: contig of 663 bp in length  
 53519 53472: gap of unknown length  
 53520 53472: contig of 659 bp in length  
 53521 53472: gap of unknown length  
 53522 53472: contig of 630 bp in length  
 53523 53472: gap of unknown length  
 53524 53472: contig of 748 bp in length  
 53525 53472: gap of unknown length  
 53526 53472: contig of 520 bp in length  
 53527 53472: gap of unknown length  
 53528 53472: contig of 679 bp in length  
 53529 53472: gap of unknown length  
 53530 53472: contig of 629 bp in length  
 53531 53472: gap of unknown length  
 53532 53472: contig of 618 bp in length  
 53533 53472: gap of unknown length  
 53534 53472: contig of 845 bp in length.

FEATURES  
 source

1. .ll5166  
 /organism="Drosophila melanogaster"  
 /strain="y; cn bw sp"

Query Match 2.6%; Score 21; DB 33; Length 115166;  
 Best Local Similarity 100.0%; Pred. No. 0.57;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 138 CCTGGCGGAGCGGAGCGG 158  
 |||||  
 Db 90770 CCTGGCGGAGCGGAGCGG 90790  
 |||||  
 RESULT 4  
 AC008231 158018 bp DNA HTG 02-AUG-1999  
 LOCUS Drosophila melanogaster chromosome 2 clone BACR12013 (D946) RPCI-98  
 DEFINITION 12.O.13 map 57B-57D strain y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 158 unsorted pieces.  
 AC008231  
 VERSION HTG; HTGS\_PHASE1  
 KEYWORDS fruit fly.  
 SOURCE Drosophila melanogaster  
 ORGANISM  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE  
 1 (bases 1 to 158018)  
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
 Butenhoff, C., Champ, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleab, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.  
 TITLE Sequencing of Drosophila melanogaster  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 158018)  
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
 Butenhoff, C., Champ, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleab, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 COMMENT  
 On Aug 2, 1999 this sequence version replaced gi:5649189.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
 the following cutoffs: length >= 200 bases.  
 \* NOTE: This is a 'working draft' sequence.  
 \* consists of 158 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 695: contig of 695 bp in length  
 \* 696: gap of unknown length  
 \* 775: contig of 682 bp in length  
 \* 776: gap of unknown length  
 \* 1458: contig of 576 bp in length  
 \* 1538: gap of unknown length  
 \* 2114: contig of 873 bp in length  
 \* 2193: gap of unknown length  
 \* 3066: contig of 1089 bp in length  
 \* 3146: gap of unknown length  
 \* 3067: contig of 1121 bp in length  
 \* 3147: gap of unknown length  
 \* 4236: contig of 453 bp in length  
 \* 4316: gap of unknown length  
 \* 5437: contig of 453 bp in length  
 \* 5517: gap of unknown length

```

CDS      complement(join(23119..23276,23439..23602,23686..23953,
24029..24192,24374..24510))
        /gene="T23K8.6"
        /codon_start=1

Query Match      2.78; Score 22; DB 8; Length 53090;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 620 CTTTGGTTTCTTAAAGTTTG 641
      |||||
Db 28604 CTTTGGTTTCTTAAAGTTTG 28583

RESULT 3
LOCUS      AC007589 115166 bp DNA HTG 02-AUG-1999
DEFINITION Drosophila melanogaster chromosome 3 clone BACR20D10 (D867) RPCI-98
            *** 73 unordered pieces.
ACCESSION  AC007589
VERSION     AC007589.3 GI:5670563
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 115166)
AUTHORS    Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE       Sequencing of Drosophila melanogaster
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 115166)
AUTHORS    Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE       Direct Submission
JOURNAL     Submitted (19-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 2, 1999 this sequence version replaced gi:5649299.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence.
            * consists of 73 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1 1144: contig of 1144 bp in length
            * 1145 1224: gap of unknown length
            * 1225 2182: contig of 958 bp in length
            * 2183 2262: gap of unknown length
            * 2263 3020: contig of 758 bp in length
            * 3021 3100: gap of unknown length

```

```

3101 3657: contig of 557 bp in length
3658 3737: gap of unknown length
3738 4493: contig of 756 bp in length
4494 4573: gap of unknown length
4574 5654: contig of 1081 bp in length
5655 5734: gap of unknown length
5735 6436: contig of 702 bp in length
6437 7312: contig of 796 bp in length
7313 7392: gap of unknown length
7393 8314: contig of 922 bp in length
8315 8394: gap of unknown length
8395 9075: contig of 681 bp in length
9076 9156: gap of unknown length
9157 9709: contig of 554 bp in length
9710 9789: gap of unknown length
9790 10520: contig of 731 bp in length
10521 10600: gap of unknown length
10601 11335: contig of 735 bp in length
11336 11415: gap of unknown length
11416 12536: contig of 1121 bp in length
12537 12617: gap of unknown length
12618 13730: contig of 1114 bp in length
13731 13811: gap of unknown length
13812 14642: contig of 832 bp in length
14643 14722: gap of unknown length
14723 16189: contig of 1467 bp in length
16190 16269: gap of unknown length
16270 17095: contig of 826 bp in length
17096 17175: gap of unknown length
17176 17975: contig of 800 bp in length
17976 18055: gap of unknown length
18056 18771: contig of 716 bp in length
18772 18851: gap of unknown length
18852 19711: contig of 860 bp in length
19712 19791: gap of unknown length
19792 20847: contig of 1056 bp in length
20848 20927: gap of unknown length
20928 22383: contig of 1456 bp in length
22384 22463: gap of unknown length
22464 23584: contig of 1121 bp in length
23585 23664: gap of unknown length
23665 25143: contig of 1479 bp in length
25144 25223: gap of unknown length
25224 27013: contig of 1790 bp in length
27014 27093: gap of unknown length
27094 28756: contig of 1663 bp in length
28757 28837: gap of unknown length
28838 30846: contig of 2010 bp in length
30847 30926: gap of unknown length
30927 32440: contig of 1514 bp in length
32441 32520: gap of unknown length
32521 34270: contig of 1750 bp in length
34271 34350: gap of unknown length
34351 36321: contig of 1881 bp in length
36322 36311: gap of unknown length
36312 38409: contig of 2098 bp in length
38410 38489: gap of unknown length
38490 40043: contig of 1554 bp in length
40044 40123: gap of unknown length
40124 41333: contig of 1210 bp in length
41334 41413: gap of unknown length
41414 43068: contig of 1655 bp in length
43069 43148: gap of unknown length
43149 44227: contig of 1079 bp in length
44228 44307: gap of unknown length
44308 45419: contig of 1112 bp in length
45420 45499: gap of unknown length
45500 47288: contig of 1789 bp in length
47289 47368: gap of unknown length
47369 49362: contig of 1994 bp in length
49363 49442: gap of unknown length
49443 51508: contig of 2066 bp in length

```

FEATURES		Location/Qualifiers	
source	gene	1. 53090	
		/organism="Arabidopsis thaliana"	
		/cultivar="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="1"	
gene	CDS	/clone="T23K8"	
		complement(<1..1608)	
		/gene="T23K8.1"	
		complement(join(<1..146,426..575,907..974,1044..1608))	
		/gene="T23K8.1"	
misc_feature	gene	/codon_start=1	
		/evidence=not_experimental	
		/product="T23K8.1"	
		/protein_id="AAD28867.1"	
		/db_xref="GI:4646194"	
CDS	gene	/translation="MEEQNSSEDRKSWINSKAVSGVKKTLTAGVWSSAPLLVPSL FVASTLAFSLVSPFCLFLANVACTOKVMSTLLPDTTEGGVGKDDSDGSPDEYSKIG HGEAGAGGEALPRGKEEPIPIOVKDEEMAKESTSLEKIRDGRDKEISERTIQ DDKSGNAKSEVQEPKRAPETRRGETGATKIEISTGDKDEEISNESEPIDAQS AQGTGEARKNTTKKKTKRAGNRFKLCHVNGYLVWRAAYPFVYSEQLWETMETLR KVGVSVARSATCAELKALXVFTGVEPPRSS"	
		complement(2911..7282)	
		/gene="T23K8.2"	
		complement(join(2911..3060,3150..3239,3323..3456, 3551..3696,3779..3985,4068..4265,4313..4473,4562..4717, 4797..4874,4919..5063,5159..5287,5364..6806,6842..7282))	
		/gene="T23K8.2"	
note	gene	/note="Contains similarity to F13P17.8 g1 3337355 hypothetical protein from Arabidopsis thaliana BAC gb AC004481. EST gb AA394560 comes from this gene."	
		/codon_start=1	
		/evidence=not_experimental	
		/protein_id="AAD26868.1"	
		/db_xref="GI:4646195"	
translation	gene	/translation="MKSSPRKLAQYYEGQDHKALEVIEDSILKIGKLLPDLCP LQGHFMALARTENDSDKFAFFLASVEICLEDYMLQAFVLSHLERGLRSTLYIK KAVIAKLSVMSGDPKSDQOKTKRTLEDIVETGSRIRHGLSEIKRISDSTYK AAKRLTYWAGMSAEKRNKMKVSTVELRSYVRYREGDLALQVLDASINRWK FWMRTCSOTFFYKFKHLEQVHDAKYKPVREDLAOSIDVWAGMISVADWEPDA LAAEMIKNRLEFVEFYVNGWSKDWPLAADERKLLKEIRLLVLLERKILSCS IRDMWRFPVGHLLAQFVSENYITTECLRVETPOSICFLECEDELNQILDLKLRKCE RDGTELLSMATDSUWRQVKEKIDHVCUSMLLRDLKRLKGTASDFDEGSDIVCD HNHYAKWQIICDYDECLDARNLCIQEDSRMNPEDQONIVASLLCDSCEHQTIDT ATRYAKWQIICDYDECLDARNLCIQEDSRMNPEDQONIVASLLCDSCEHQTIDT EDPLTEYLCAVRDLGIESHPDFTNAEDCLELHGLKNIISDVIKSIDLKLVS VTSKVLADSILLVENSRIHLNDRILRSVDFRSYILPLKRLRELEGIWMDA KAKLAABELLITEKKRKKSGPKKKHRSKNVAEFVRMNPWCSREHQOASVIL IRMTHTKRVLMNRESSINLEPGVSLKMWDEDSLEPERGRQETSSNTNHEEAIKD LKNPTKDSLESDATRYRSALDMLKNSPTKIRFKDLSPLLIQIRYFTALLNI KVLQEDLVNRPQPHGNLEEQVPYALQNLFAISVEQIAEGLYSYLLSNLASLEGV HMSDSDADVVAAILLEPFCNKSPPERSLVTLEFLAERYMSCKKCKRKNYPQSS YGIYMAANSIRNLKCAFNGIKFEDILKMRKDMICDVKGGCGDINFVHTISRCP PIFTVLBWNETEKDISETKALDWEIDISRYVEGLEESNIRKRLVSMVSGCVGE YICWAYKKNRWSURHEALAEVVGNNVNVVRCGERKVRPEILFYEAQFQPNKWQKL ISRKYQVENLAFM"	
		9509..9875	
		/note="Homologous to u5 small nuclear RNA gb X13012."	
		complement(10409..14859)	
		/gene="T23K8.3"	
CDS	gene	complement(join(10409..10498,10597..10686,10775..10908, 11005..11152,11241..11447,11532..11729,11753..11958, 12040..12229,12300..13229,12416..12550,12633..12761, 12835..14859))	
		/gene="T23K8.3"	
		/note="Contains similarity to F13P17.8 g1 3337355 hypothetical protein from Arabidopsis thaliana BAC gb AC004481."	
		/codon_start=1	
		/evidence=not_experimental	
translation	gene	/protein_id="AAD26881.1"	
		/db_xref="GI:4646208"	
		/translation="MAPLLPLLVDPPEAKIQKLHGVFAEGDIKALEVIEDSISTR ILKVSQNSTIQNESVLSWNSVFSVLSMDYVQEAQNSLTDNESPLQDILIRFMLNN LSLTHPLEANGYSKFLNILELIPRNHYFVNEVAKTICIRCKTDMAYSGEYSYGIIL INANSLRVYKSAFKEFTFENIIKASINVMKCDCKEGCKRNVYDMISNLPASIVYA LQWNETEKEIFDASVLATIEDISAIYRVGDSAFTRYRLVSMVSHGNLCNCAVY ENNRVVRHFCSEIEVIGDWGVLRIFRELHIRPEILFFENAILKQDMFSQSGED"	
		complement(23119..24510)	
		/gene="T23K8.6"	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 199723)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 199723)  
Waterston,R.H.  
Direct Submission  
Submitted (04-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

Center project name: H.NH0521M14.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
2353: contig of 2353 bp in length  
2354  
2371: gap of unknown length  
2372  
4340: contig of 1969 bp in length  
4341  
4358: gap of unknown length  
4359  
6425: contig of 2067 bp in length  
6426  
6443: gap of unknown length  
6444  
9090: contig of 2647 bp in length  
9091  
9108: gap of unknown length  
9109  
12172: contig of 3064 bp in length  
12173  
12190: gap of unknown length  
12191  
16034: contig of 3844 bp in length  
16035  
16052: gap of unknown length  
16053  
19152: contig of 3100 bp in length  
19153  
19170: gap of unknown length  
19171  
22944: contig of 3774 bp in length  
22945  
22962: gap of unknown length  
22963  
27695: contig of 4733 bp in length  
27696  
27713: gap of unknown length  
27714  
31855: contig of 4142 bp in length  
31856  
31873: gap of unknown length  
31874  
35780: contig of 3907 bp in length  
35781  
35798: gap of unknown length  
35799  
40716: contig of 4918 bp in length  
40717  
40734: gap of unknown length  
40735  
45236: contig of 4502 bp in length  
45237  
45254: gap of unknown length  
45255  
48959: contig of 3705 bp in length  
48960  
48977: gap of unknown length  
48978  
55337: contig of 6360 bp in length  
55338  
55355: gap of unknown length  
55356  
62799: contig of 7444 bp in length  
62800  
62817: gap of unknown length  
62818  
70545: contig of 7728 bp in length  
70546  
70563: gap of unknown length  
70564  
77953: contig of 7390 bp in length  
77954  
77971: gap of unknown length  
77972  
84023: contig of 6052 bp in length  
84024  
84041: gap of unknown length  
84042  
90480: contig of 6439 bp in length  
90481  
90498: gap of unknown length  
90499  
97338: contig of 6840 bp in length  
97339  
97356: gap of unknown length  
97357  
108436: contig of 11080 bp in length  
108437  
108454: gap of unknown length  
108455  
116696: contig of 8242 bp in length  
116697  
116714: gap of unknown length  
116715  
124591: contig of 7877 bp in length  
124592  
124609: gap of unknown length  
124610  
135616: contig of 11007 bp in length  
135617  
135634: gap of unknown length  
135635  
147194: contig of 11560 bp in length  
147195  
147211: gap of unknown length  
147212  
159038: contig of 11827 bp in length  
159039  
159055: gap of unknown length

\* 159056 178182: contig of 19127 bp in length  
\* 178183 178199: gap of unknown length  
\* 178200 199723: contig of 21524 bp in length.  
FEATURES  
Location/Qualifiers  
1. 199723  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-521M14"  
BASE COUNT 57984 a 40505 c 40199 g 60534 t 501 others  
ORIGIN

Query Match 2.8%; Score 23; DB 44; Length 199723;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 606 CTTGGATACCTGCTTTGGGTT 628  
|||||  
Db 102200 CTTGGATACCTGCTTTGGGTT 102222

RESULT 2  
T23K8/c  
LOCUS T23K8 53090 bp DNA PLN 13-MAY-1999  
DEFINITION Arabidopsis thaliana chromosome 1 BAC T23K8 sequence, complete  
sequence.  
ACCESSION AC007230.2 GI:4582411  
VERSION AC007230  
KEYWORDS HTG.  
SOURCE  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;  
Arabidopsis.  
1 (bases 1 to 53090)  
Vysotskaia,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,  
Li,J., Kremenetskaia,I., Ngan,I., Luros,J., Gonzalez,A., Altafi,H.,  
Araujo,R., Chao,Q., Conn,L., Conway,A.B., Dunn,P., Hansen,N.,  
Huizar,L., Kim,C., Palm,C.J., Rowley,D., Shinn,P., Walker,M.,  
Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.  
Arabidopsis thaliana chromosome 1 BAC T23K8 sequence  
Unpublished (1999)  
This sequence is of BAC T23K8 from Arabidopsis thaliana chromosome  
1. The sequence does not represent the sequence of the entire  
insert of this clone. It is shorter by 31564 bp because we submit  
only the unique sequence of the clone. However, in order to  
facilitate the joining of overlapping clones in the future for  
creation of larger contigs, we provide a small overlap (200 bp)  
between overlapping submitted clones. The 3' end of this sequence  
overlaps by 200 bp the 5' end of the sequence of the BAC T8F5.  
2 (bases 1 to 53090)  
Theologis,A.  
Direct Submission  
Submitted (06-APR-1999) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
3 (bases 1 to 53090)  
Theologis,A.  
Direct Submission  
Submitted (13-APR-1999) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
4 (bases 1 to 53090)  
Theologis.  
Direct Submission  
Submitted (22-APR-1999) Plant Gene Expression Center, 800 Buchanan  
St., Albany, CA 94710, USA  
5 (bases 1 to 53090)  
Theologis.  
Direct Submission  
Submitted (13-MAY-1999) Plant Gene Expression Center, 800 Buchanan  
St., Albany, CA 94710, USA  
On Apr 13, 1999 this sequence version replaced gi:4567183  
The sequence of BAC T23K8 from Arabidopsis thaliana chromosome 1.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1. The sequence does not represent the sequence of the entire  
insert of this clone. It is shorter by 31564 bp because we submit  
only the unique sequence of the clone. However, in order to  
facilitate the joining of overlapping clones in the future for  
creation of larger contigs, we provide a small overlap (200 bp)  
between overlapping submitted clones. The 3' end of this sequence  
overlaps by 200 bp the 5' end of the sequence of the BAC T8F5.  
2 (bases 1 to 53090)  
Theologis,A.  
Direct Submission  
Submitted (06-APR-1999) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
3 (bases 1 to 53090)  
Theologis,A.  
Direct Submission  
Submitted (13-APR-1999) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
4 (bases 1 to 53090)  
Theologis.  
Direct Submission  
Submitted (22-APR-1999) Plant Gene Expression Center, 800 Buchanan  
St., Albany, CA 94710, USA  
5 (bases 1 to 53090)  
Theologis.  
Direct Submission  
Submitted (13-MAY-1999) Plant Gene Expression Center, 800 Buchanan  
St., Albany, CA 94710, USA  
On Apr 13, 1999 this sequence version replaced gi:4567183  
The sequence of BAC T23K8 from Arabidopsis thaliana chromosome 1.

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 05:25:14 ; Search time 553.5 Seconds  
(without alignments)  
-4443.489 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTTAAAGGTC.....AAACGAATTATATCCATAA 810

Scoring table: OLIGO\_NUC  
Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl :  
Word size : 0  
Number of hits that pass the threshold : 1642386

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pl1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_v1.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_v1.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*
- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*

50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	2.8	199723	44	AC016756	AC016756 Homo sapi
2	22	2.7	53090	8	T23R8	AC007230 Arabidops
3	21	2.6	115166	3	AC007589	AC007589 Drosophil
4	21	2.6	158018	33	AC008231	AC008231 Drosophil
5	21	2.6	3937	34	DMU35074	U35074 Drosophila
6	21	2.6	113562	42	AC011072	AC011072 Drosophil
7	21	2.6	76396	43	AC012895	AC012895 Drosophil
8	21	2.6	27387	45	AC017975	AC017975 Drosophil
9	20	2.5	100000	10	AP000076	AP000076 Homo sapi
10	20	2.5	98209	11	HS260J9	282193 Human DNA s
11	20	2.5	62216	44	AC016414	AC016414 Homo sapi
12	20	2.5	67795	44	AC016538	AC016538 Homo sapi
13	20	2.5	60013	44	AC016814	AC016814 Mus muscu
14	19	2.3	1851	4	CAVHH6X	X65266 C.auratus i
15	19	2.3	2069	8	AF194825	AF194825 Bougainvi
16	19	2.3	89137	8	ATAC006284	AC006284 Arabidops
17	19	2.3	194430	10	CNS01DS2	AL121652 BAC seque
18	19	2.3	171556	10	CNS01DS8	AL121658 BAC seque
19	19	2.3	220895	10	HS250D10	299716 Human DNA s
20	19	2.3	2124	10	HS801125	AL117592 Homo sapi
21	19	2.3	126963	11	HS753D4	AL031676 Human DNA
22	19	2.3	16437	14	AF123770	AF123770 Shuttle v
23	19	2.3	18772	14	AF133437	AF133437 Cloning v
24	19	2.3	16297	14	CVU75991	U75991 Cloning vec
25	19	2.3	19500	14	CVU75992	U75992 Cloning vec
26	19	2.3	11612	14	CVU80929	U80929 Cloning vec
27	19	2.3	8279	19	HS131497	A131497 Homo Sapi
28	19	2.3	53517	32	AC002405	AC002405 Homo sapi
29	19	2.3	197010	32	AL133159	AL133159 Mus muscu
30	19	2.3	197010	32	AL133160	AL133160 Mus muscu
31	19	2.3	151245	32	AL133230	AL133230 Homo sapi
32	19	2.3	151210	32	AP000457	AP000457 Homo sapi
33	19	2.3	105079	32	AP000632	AP000632 Homo sapi
34	19	2.3	176875	32	CNS0000K	AL049840 Homo sapi
35	19	2.3	206273	32	CNS0000S	AL049870 Homo sapi
36	19	2.3	190880	32	CNS000M8	AL079307 Homo sapi
37	19	2.3	177617	32	CNS00M8T	AL079343 Homo sapi
38	19	2.3	216550	32	CNS00M8U	AL079351 Homo sapi
39	19	2.3	240250	32	CNS00YVG	AL096821 Homo sapi
40	19	2.3	183059	32	CNS01DRP	AL117692 Homo sapi
41	19	2.3	201658	32	CNS01DRX	AL121579 Homo sapi
42	19	2.3	243962	32	CNS01DRY	AL121594 Homo sapi
43	19	2.3	216986	32	CNS01DS1	AL121612 Homo sapi
44	19	2.3	194957	32	CNS01DTI	AL132777 Homo sapi
45	19	2.3	176667	45	AC016927	AC016927 Homo sapi

ALIGNMENTS

RESULT 1

AC016756  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC016756 199723 bp DNA HTG  
Homo sapiens clone Rpl1-521M14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 29  
unordered pieces.

AC016756.1 GI:6524288  
HTG; HTGS\_PHASE1.  
human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
Query Match      2.1%; Score 17; DB 1; Length 2625;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 335 ACCTCAAAGTGGACAA 351  
|||||  
Db 1223 ACCTCAAAGTGGACAA 1239

RESULT 1.5

T31984  
ID T31984 standard; DNA; 8075 BP.  
AC T31984;  
AD 16-JAN-1997 (first entry)  
DE Invertebrate calcium channel subunit coding sequence.  
DE DE Invertebrate calcium channel subunit; pharmacology; receptor;  
KW screening; identification; drug; insecticide;  
KW Drosophila melanogaster; ss.  
Drosophila melanogaster.  
Key Location/Qualifiers  
FT 157..7708  
ET cds  
FT /\*tag= a  
FT /product= Calcium channel subunit  
PN W09622307-nl.  
PD 25-JUL-1996.  
PP 19-JAN-1996; U00887.  
PR 19-JAN-1995; US-374888.  
PY (UNYK ) UNIV NEW YORK STATE RES FOUND.  
PI Dubald MM, Hall LM, Ren D, Zheng W;  
DR WPI: 96-354475/35.  
PT P-PSDB: W01884.  
PT Gene encoding invertebrate calcium channel subunit - useful for  
DR screening cpds. for activity as pesticides and drugs  
PT Claim 7; Page 70-80; 149pp; English.  
PT The gene encoding the neuronal invertebrate calcium channel  
CCC subunit can be used to transform host cells which then express  
CCC functional calcium channels with the same pharmacological properties  
CCC as receptors in neuronal invertebrate tissue. Cells carrying this  
CCC calcium channel can be used to identify insecticides (especially  
CCC those having no harmful effects on vertebrates) and drugs.  
SQ Sequence 8075 BP; 2093 A; 1935 C; 2081 G; 1966 T;

```
Query Match      2.1%; Score 17; DB 1; Length 8075;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

148 GC CGGAAGCGGAGTGG 164  
|||||  
Db 7432 GC CGGAAGCGGAGTGG 7448

Search completed: February 11, 2000, 03:37:01  
Job time: 211 sec



```
OS Acremonium chrysogenum.
FH Key Location/Qualifiers
FT cds 1252..2717
FT /*tag= a
FT /product= PGK
FT intron 1318..1462
FT /*tag= b
FT intron 1884..1947
FT /*tag= c
FT J05213998-A.
FT 24-AUG-1993.
FT 02-AUG-1991. 282369.
FT 03-AUG-1990; JP-204978.
FT 30-JUL-1991; JP-189984.
FT (ASAH ) ASAH CHEM IND CO LTD.
FT WPI; 93-299652/28.
FT P-PSDB; 849247.
FT Novel polypeptide obtd. by culturing transformed fungus - having
FT blood coagulation preventing, platelet aggregation preventing and
FT thrombolytic activities
FT PGK is described in Reference Example 2 (1).
FT Novel polypeptides, obtd. by culturing transformed fungus, have
FT blood coagulation preventing, platelet aggregation preventing
FT and thrombolytic activities.
FT In an example, plasmid M13mp19TMD3 (constructed from pSV2TMD2
FT (ATCC 67238) contg. a human thrombomodulin sequence) was subjected
FT to site directed mutagenesis to prepare pSV2TMD7. Plasmid pSV2TMD7
FT was transfected to COS-1 cells. The activity of promoting protein C
FT activation by thrombin of the peptide produced by the transformed
FT COS-1 cell was measured. The amt. of the peptide was determined.
FT Sequence 3306 BP; 787 A; 1055 C; 885 G; 579 T;
SQ Sequence 3306 BP; 787 A; 1055 C; 885 G; 579 T;

Query Match 2.18; Score 17; DB 1; Length 3306;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 CTCCTCCCGCTCTCTGG 142
Db 209 CTCCTCCCGCTCTCTGG 193

RESULT 13
O73500
ID Q73500 standard; DNA; 8438 BP.
DE DNA encoding Pseudorabies virus large latency transcript.
KW Pseudorabies virus; PRV; LLV; large latency transcript;
KW attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICP0;
KW protecting animals; deletion mutants; swine; ds.
OS Pseudorabies virus.
FH Key Location/Qualifiers
FT misc_feature 1..7013
FT /*tag= a
FT /note= "derived from PRV strain InFh"
FT misc_feature 7014..8425
FT /*tag= b
FT /note= "derived from PRV strain Ka"
FT cds 622..6498
FT /*tag= c
FT /note= "encodes predicted amino acid sequence of ORF2"
FT tata_signal 1..6
FT /*tag= d
FT misc_feature 34
FT /*tag= e
FT /note= "RNA cap site"
FT polya_signal 8382..8387
FT /*tag= f
FT US5352596-A.
FT 04-OCT-1994.
FT 11-SEP-1992; 945283.

PR 11-SEP-1992; US-945283.
PA (USDA ) US SEC OF AGRIC.
PI Cheung AK, Wesley RD;
DR WPI; 94-316187/39.
DR P-PSDB; R60620.
PT New pseudorabies virus mutants for use in vaccine - having a
PT deletion and/or insertion in the early protein O gene or large
PT latency transcript gene
PS Disclosure; Column 15-30; 43pp; English.
CC Q73500 shows the Pseudorabies virus (PRV) large latency transcript
CC (LLT). The basic sequence is derived from PRV strain InFh and PRV
CC strain Ka. The LLT overlaps and is transcribed in the opposite
CC orientation with respect to the EP0 (early polypeptide 0) and the
CC immediately early gene (IE180). EP0 is nonessential for replicatio,
CC LLT is the only gene expressed during PRV latency, and the IE180
CC gene is absolutely necessary for PRV replication. However there are
CC 2 copies of IE180 in the genome. It is expected that PRV lacking one
CC of the IE180 copies is viable. Deletions in the non-overlapping
CC regions of these 3 genes will generate single deletion routants,
CC while deletions in overlapping regions will generate double deletion
CC mutants. The invention is concerned with the construction of attenuated
CC viruses which have a reduced ability to reactivate from latency. This
CC can be achieved by functionally disabling the expression of the EP0
CC gene, or by disrupting the synthesis of the LLT, or both.(See also
SQ Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T;

Query Match 2.18; Score 17; DB 1; Length 8438;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 182 CCAAGATGGCGCGGCC 198
Db 7222 CCAAGATGGCGCGGCC 7238

RESULT 14
O87970
ID Q87970 standard; cDNA; 2625 BP.
AC Q87970;
DT 06-NOV-1995 (first entry)
DE Peptidyl C-terminal alpha-amidating enzyme cDNA.
KW Peptidyl C-terminal alpha-amidating enzyme; AE; cell culture;
KW trichostatin; CHO; ds.
OS Not specified.
FH Key Location/Qualifiers
FT cds 1..2625
FT /*tag= a
FT signal_peptide 1..117
FT /*tag= b
FT mat_peptide 118..2625
FT /*tag= c
FT EP-649900-A.
FT 26-APR-1995.
FT 07-SEP-1994; 306587.
FT 08-SEP-1993; JP-257881.
FT (FURU/) FURUKAWA K.
FT (SUNR ) SUNTORY LTD.
FT Furukawa K, Ohsuye K, Sugimura K;
FT WPI; 95-156754/21.
FT P-PSDB; R73053.
PT Increasing protein prodn. from cultured animal cells - by adding
PT a trichostatin to the medium, effective at low concn. and not
PT injurious to host cells
PS Disclosure; Page 10-15; 19pp; English.
CC 3mu-1S cells (i.e. CHO cells expressing a peptidyl C-terminal alpha-
CC amidase enzyme (AE)) were suspended in F-12 medium to which aliquots
CC of trichostatin were added. Cells were cultured for 3 days at 37 deg
CC and then assayed for AE. Without trichostatin, AE productivity was
CC 708 U/ml. At 10, 50, 100 and 200 nM trichostatin, AE productivities
CC were respectively 866, 1897, 1894 and 3359 U/ml.
SQ Sequence 2625 BP; 761 A; 563 C; 633 G; 668 T;
```

DR WPI: 92-010570/02.  
DR P-PSDB: R20112.  
PT Novel DNA encoding peptidyl hydroxyglycine N-C lyase (PHL) -  
PT used to prepare PHL which can be used in the amidation of  
PT peptide(s) e.g. human calcitonin.  
PS Claim 9; Page 18; 28pp; English.  
CC The sequence represents an insert from pAE-III-202-4 (FERM BP-3172).  
CC The plasmid was prep'd. by ligation of fragments from pAE-II-202-1  
CC and pAE-III-202-2 (see feature table), two subclones of AE-III-202,  
CC a clone isolated from a cDNA library in lambda gt10. The library  
CC as prep'd. from RNA extracted from the skin of two heads of Xenopus  
CC laevis. The vector serves as a source for a DNA fragment encoding  
CC PHL for the construction of an expression vector for the prep'n. of  
CC recombinant PHL. The PHL catalyses the reaction: R-GlyOH -> R-NH2.  
CC It can be used to produce peptides with amidated C-termini, e.g.  
CC calcitonin, growth hormone, LH-RH.  
SQ Sequence 3383 BP; 970 A; 702 C; 795 G; 916 T;

Query Match 2.1%; Score 17; DB 1; Length 3383;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAAGTGGGACAA 351  
Db 1241 ACCTCAAGTGGGACAA 1257  
|||||

RESULT 10  
Q22202/C  
ID Q22202 standard; DNA: 3306 BP.  
AC Q22202;  
DT 03-JUL-1992 (first entry)  
DE A. chrysoygenum phosphoglycerate kinase.  
KW Fungal PGK; yeast; ss.  
OS Acromonium chrysoygenum.

FT key location/Qualifiers  
FT exon 1252..1317  
FT FT /\*tag= a  
FT FT /number= 1  
FT FT 1318..1462  
FT FT /\*tag= b  
FT FT /number= 1  
FT FT 1463..1883  
FT FT /\*tag= c  
FT FT /number= 2  
FT FT 1884..1947  
FT FT /\*tag= d  
FT FT /number= 2  
FT FT 1948..2717  
FT FT /\*tag= e  
FT FT /number= 3

PN EP-474273-A.  
PD 11-MAR-1992.  
PF 05-AUG-1991; 202009.  
PR 03-AUG-1990; JP-204978.  
PA (ASAH ) ASAH KASEI KOGYO.  
PI Zushi M, Gomi K, Yamamoto S, Suzuki K, Matsuda A;  
DR WPI: 92-081820/11.  
DR P-PSDB: R22025.  
PT New polypeptide inhibits blood coagulation and platelet  
PT aggregation - promotes thrombin catalysed protein C activation  
PT for treating myocardial infarction, thrombosis, embolism, etc.  
PS Example 2(Ref); Fig 61; 112pp; English.  
CC The phosphoglycerate kinase gene was isolated from a total DNA  
CC of A.chrysoygenum cloned in a Lambda library. Its isolation and  
CC sequencing are described in Reference example 2-(1).  
SQ Sequence 3306 BP; 787 A; 1042 C; 898 G; 579 T;

Query Match 2.1%; Score 17; DB 1; Length 3306;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CTCCCTCCCGCTCCTGG 142  
Db 209 CTCCCTCCCGCTCCTGG 193  
|||||

RESULT 11  
Q23005/C  
ID Q23005 standard; DNA: 3306 BP.  
AC Q23005;  
DT 31-JUL-1992 (first entry)  
DE Phosphoglycerate kinase gene (incl. promoter).  
KW PGK; fermentation; cephalosporin; promoter; ss.  
OS Acromonium chrysoygenum.  
FT key location/Qualifiers  
FT promoter 1..1251  
FT FT /\*tag= a  
FT FT /note= "claim 1, page 1"  
FT FT 1252..2717  
FT FT /\*tag= b  
FT FT /product= PGK  
FT FT 1252..1317  
FT FT /\*tag= c  
FT FT /number= 1  
FT FT 1318..1462  
FT FT /\*tag= d  
FT FT /number= 1  
FT FT 1463..1883  
FT FT /\*tag= e  
FT FT /number= 2  
FT FT 1884..1947  
FT FT /\*tag= f  
FT FT /number= 2  
FT FT 1948..2114  
FT FT /\*tag= g  
FT FT /number= 3

PN J04058891-A.  
PD 25-FEB-1992.  
PF 27-JUN-1990; 166566.  
PR 27-JUN-1990; JP-166566.  
PA (ASAH ) ASAH CHEMICAL IND KK.  
DR WPI: 92-111675/14.  
DR P-PSDB: R22095.  
PT Acromonium chrysoygenum promoter for efficient gene expression -  
PT comprises DNA fragment having promoter active part of  
PT phospho-glycerate kinase gene of Acromonium chrysoygenum, etc.  
PS Disclosure; Fig 4; 18pp; Japanese.  
CC Various genes can be expressed efficiently in A. chrysoygenum  
CC using the promoter indicated in the features. The fermentation  
CC yield of cephalosporin can be improved using vectors contg. the  
CC promoter.  
SQ Sequence 3306 BP; 787 A; 1041 C; 899 G; 579 T;

Query Match 2.1%; Score 17; DB 1; Length 3306;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CTCCCTCCCGCTCCTGG 142  
Db 209 CTCCCTCCCGCTCCTGG 193  
|||||

RESULT 12  
Q48534/C  
ID Q48534 standard; DNA: 3306 BP.  
AC Q48534;  
DT 30-MAR-1994 (first entry)  
DE PGK.

KW Transformation; fungus; blood coagulation; prevention; platelet; PGK;  
KW aggregation; thrombolytic activity; thrombolysis; human thrombomodulin;  
KW site-directed mutagenesis; promotion; protein C; activation;  
KW thrombin; ss.

ID N92871 standard; DNA; 1665 BP.  
AC N92871;  
DT 27-JUN-1990 (first entry)  
DE Expression plasmid pUCpLCI799 RV.  
KW alpha-amidating; pAX799; alpha amide; ds.  
OS Synthetic.  
PN EP-299790-A.  
PD 18-JAN-1989.  
PF 15-JUL-1988; 306508.  
PR 17-JUL-1987; JP-177184.  
PR 05-DEC-1987; JP-306867.  
PA (SUNR) Sutory Ltd.  
PI Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;  
DR WPI; 89-017279/03.  
PT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -  
and their precursors deoxyribonucleic acid encoding sequences.  
PS Disclosure; 7pp; English.  
CC The sequence encodes a derivative of the mature C-terminal alpha-  
amidating enzyme from plasmid pXA799.  
CC The plasmid was screened from an E.coli library using plasmid pXA457  
to screen a larger library.  
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of  
hydrophobic elements suggesting a membrane function.  
CC See also N93060.  
SQ Sequence 1665 BP; 480 A; 367 C; 393 G; 425 T;

Query Match 2.1%; Score 17; DB 1; Length 1665;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAAGTGGGACAA 351  
|||||  
DB 1106 ACCTCAAGTGGGACAA 1122

## RESULT 7

ID N92872 standard; DNA; 1518 BP.  
AC N92872;  
DT 27-JUN-1990 (first entry)  
DE Expression plasmid pUCpLCI799 Sali.  
KW alpha-amidating; pAX799; alpha amide; ds.  
OS Synthetic.  
PN EP-299790-A.  
PD 18-JAN-1989.  
PF 15-JUL-1988; 306508.  
PR 17-JUL-1987; JP-177184.  
PR 05-DEC-1987; JP-306867.  
PA (SUNR) Sutory Ltd.  
PI Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;  
DR WPI; 89-017279/03.  
PT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -  
and their precursors deoxyribonucleic acid encoding sequences.  
PS Disclosure; 7pp; English.  
CC The sequence encodes a derivative of the mature C-terminal alpha-  
amidating enzyme from plasmid pXA799.  
CC The plasmid was screened from an E.coli library using plasmid pXA457  
to screen a larger library.  
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of  
hydrophobic elements suggesting a membrane function.  
CC See also N93060.  
SQ Sequence 1518 BP; 445 A; 321 C; 361 G; 391 T;

Query Match 2.1%; Score 17; DB 1; Length 1518;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAAGTGGGACAA 351  
|||||

Db 1106 ACCTCAAGTGGGACAA 1122

## RESULT 8

ID N94527 standard; DNA; 2079 BP.  
AC N94527;  
DT 27-JUN-1990 (first entry)  
DE Expression plasmid pUCpLCI799 BglII.  
KW alpha-amidating; pAX799; alpha amide; ds.  
OS Synthetic.  
PN EP-299790-A.  
PD 18-JAN-1989.  
PF 15-JUL-1988; 306508.  
PR 17-JUL-1987; JP-177184.  
PR 05-DEC-1987; JP-306867.  
PA (SUNR) Sutory Ltd.  
PI Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;  
DR WPI; 89-017279/03.  
PT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -  
and their precursors deoxyribonucleic acid encoding sequences.  
PS Disclosure; 7pp; English.  
CC The sequence encodes a derivative of the mature C-terminal alpha-  
amidating enzyme from plasmid pXA799.  
CC The plasmid was screened from an E.coli library using plasmid pXA457  
to screen a larger library.  
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of  
hydrophobic elements suggesting a membrane function.  
CC See also N93060.  
SQ Sequence 2079 BP; 606 A; 445 C; 500 G; 528 T;

Query Match 2.1%; Score 17; DB 1; Length 2079;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 ACCTCAAGTGGGACAA 351  
|||||  
DB 1106 ACCTCAAGTGGGACAA 1122

## RESULT 9

ID Q20269 standard; CDNA; 3383 BP.  
AC Q20269;  
DT 06-APR-1992 (first entry)  
DE Gene encoding AE-III (peptidylhydroxyglycine N-C lyase precursor).  
KW Amidation; PHL; ds.  
OS Xenopus laevis.  
FH Key Location/Qualifiers  
FT cds 31..2838  
FT /tag= a  
FT /product= AE-III  
FT /note= "including PAM and PHL domains"  
FT mat\_peptide 1177..2145  
FT /tag= b  
FT /product= PHL  
FT misc\_feature 1..1891  
FT /tag= c  
FT /note= "from pAE-III-202-1"  
FT misc\_feature 1891..3383  
FT /tag= d  
FT /note= "from pAE-III-202-2"  
PN EP-465404-A.  
PD 08-JAN-1992.  
PF 27-MAY-1991; 810399.  
PR 01-JUN-1990; JP-141678.  
PR 10-AUG-1990; JP-210335.  
PR 30-NOV-1990; JP-329911.  
PA (CIBA ) CIBA GEIGY AG.  
PI Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y,  
Kawahara T, Kangawa K;



241 CTCGTTGTCCTGTCGTTGTCCTCACTACAGACCCCTGGGGGCTGTTCCACC 300  
|||||  
241 CTCGTTGTCCTGTCGTTGTCCTCACTACAGACCCCTGGGGGCTGTTCCACC 300  
|||||  
301 TCCGCGGGGGGAGGAGTGCCTTAAGTCGAGGACCTCAAGTGGGCAATATATTGT 360  
|||||  
301 TCCGCGGGGGGAGGAGTGCCTTAAGTCGAGGACCTCAAGTGGGCAATATATTGT 360  
|||||  
361 AAAGATCCAAAATAAATAGCGCTACGCAAGAACAGTAACTGTACAACTACACAGCT 420  
|||||  
361 AAAGATCCAAAATAAATAGCGCTACGCAAGAACAGTAACTGTACAACTACACAGCT 420  
|||||  
421 CATGTTCTCTGTTTCCAGCACCCCAACATACTGTAAGGATTCAGTGGCAATGAACA 480  
|||||  
421 CATGTTCTCTGTTTCCAGCACCCCAACATACTGTAAGGATTCAGTGGCAATGAACA 480  
|||||  
481 CATTTTACTGGGAAGAGTGGTGTGTTTCAAGCCCATCTTGCAGAAATGAATGGC 540  
|||||  
481 CATTTTACTGGGAAGAGTGGTGTGTTTCAAGCCCATCTTGCAGAAATGAATGGC 540  
|||||  
541 TATTCCTACAAAGTGGCAGTGCATGCTCTTTTCTGATGGTGGAGCAGATCGA 600  
|||||  
541 TATTCCTACAAAGTGGCAGTGCATGCTCTTTTCTGATGGTGGAGCAGATCGA 600  
|||||  
601 TTTTACCTGGATACCCCTGCTTTGGGTTTGTAAAGTTTGCACGTGTAGGGTTTGTGA 660  
|||||  
601 TTTTACCTGGATACCCCTGCTTTGGGTTTGTAAAGTTTGCACGTGTAGGGTTTGTGA 660  
|||||  
661 ATGGGAGCCCTAATGATTCATCTTATTTCAATGCGAGATGTTGGACCTTCAGATGA 720  
|||||  
661 ATGGGAGCCCTAATGATTCATCTTATTTCAATGCGAGATGTTGGACCTTCAGATGA 720  
|||||  
721 AGTAGTTACATTATAGATTACTATGGAACACAGCTTACAGACTAGTATTACTAATGAA 780  
|||||  
721 AGTAGTTACATTATAGATTACTATGGAACACAGCTTACAGACTAGTATTACTAATGAA 780  
|||||  
781 ACATTAGAAAAACGCAATTATATCCATAA 810  
|||||  
781 ACATTAGAAAAACGCAATTATATCCATAA 810  
|||||

## RESULT 2

X41191  
ID X41191 standard; cDNA; 440 BP.  
AC X41191.  
DT 17-JUN-1999 (first entry)  
Human secreted protein 5' EST SEQ ID NO:135.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN W0906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
FA (GEST) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153778/13.  
DR P-PSDB; Y12358.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 1; Page 315; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation

CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 440 BP; 107 A; 103 C; 114 G; 114 T;

Query Match 49.3%; Score 399; DB 1; Length 440;

Best Local Similarity 100.0%; Pred. No. 1.6e-197;

Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TCTGTCTCGTCTGCTCCGGAGCCGCTGAGGCCAGACTCGTTGGTCTCTGTGGTTCGTC 264  
|||||  
Db 42 TCTGTCTCGTCTGCTCCGGAGCCGCTGAGGCCAGACTCGTTGGTCTCTGTGGTTCGTC 101  
|||||  
QY 265 TCAGTCACTACAGGACCCCTGGGGGCTGTGCCACTCCCGGGGCGAGGAGTCGCTT 324  
|||||  
Db 102 TCAGTCACTACAGGACCCCTGGGGGCTGTGCCACTCCCGGGGCGAGGAGTCGCTT 161  
|||||  
QY 325 AAGTCGGAGGACCTCAAGTGGGCAATATATTTGTAAGATCCAAAATAAATGACGCT 384  
|||||  
Db 162 AAGTCGGAGGACCTCAAGTGGGCAATATATTTGTAAGATCCAAAATAAATGACGCT 221  
|||||  
QY 385 AGCAAGAACCACTTAAGTGTACAACTACAGCTCATGTTCTCTGTTTCCAGCACCC 444  
|||||  
Db 222 AGCAAGAACCACTTAAGTGTACAACTACAGCTCATGTTCTCTGTTTCCAGCACCC 281  
|||||  
QY 445 ACATAAATCTTAAGATTCAGTGGCAATGAACACATTTTACTGGGACGAGTTGGT 504  
|||||  
Db 282 ACATAAATCTTAAGATTCAGTGGCAATGAACACATTTTACTGGGACGAGTTGGT 341  
|||||  
QY 505 TTTTTCAGCCCATATCTTCCGCAATGTAATGGCTATTCTACAAAGTGGCAGTCGCA 564  
|||||  
Db 342 TTTTTCAGCCCATATCTTCCGCAATGTAATGGCTATTCTACAAAGTGGCAGTCGCA 401  
|||||  
QY 565 TTGTCTCTTTTCTTGGATGTTGGGAGCAGATCGATTT 603  
|||||  
Db 402 TTGTCTCTTTTCTTGGATGTTGGGAGCAGATCGATTT 440  
|||||

## RESULT 3

X41259  
ID X41259 standard; cDNA; 455 BP.  
AC X41259.  
DT 17-JUN-1999 (first entry)  
Human secreted protein 5' EST SEQ ID NO:203.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN W0906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
FA (GEST) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153778/13.  
DR P-PSDB; Y12426.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 1; Page 456; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 11, 2000, 03:33:30 ; Search time 35.32 Seconds  
(without alignments)  
5579.724 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTAAAGGGTC.....AAACGCAATATATCCATAA 810

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	810	100.0	810	1 X05735	Human beta-amyloid
2	399	49.3	440	1 X41191	Human secreted pro
3	284	35.1	455	1 X41239	Human secreted pro
c 4	18	2.2	386	1 X04660	Human secreted pro
5	17	2.1	3315	1 N90791	cDNA of plasmid pX
6	17	2.1	1665	1 N92871	Expression plasmid
7	17	2.1	1518	1 N92872	Expression plasmid
8	17	2.1	2079	1 N94527	Expression plasmid
9	17	2.1	3383	1 Q20269	Gene encoding AE-1
c 10	17	2.1	3306	1 Q22202	A. chrysogenum pho
c 11	17	2.1	3306	1 Q23005	phosphoglycerate k
c 12	17	2.1	3306	1 Q48534	PGK. Novel polyep
c 13	17	2.1	8438	1 Q73500	DNA encoding Pseud
14	17	2.1	2625	1 Q87970	Peptidyl C-termina
15	17	2.1	8075	1 T31984	Invertebrate calci
16	17	2.1	8075	1 T31981	Neuronal invertebr
c 17	17	2.1	20199	1 V52139	Streptococcus pneu
c 18	17	2.1	6669	1 V55039	Human H1AP-1 codin
19	17	2.1	318	1 V88049	EST clone FQ505. N
20	17	2.1	23439	1 V74349	Staphylococcus aur
21	17	2.1	11307	1 X12986	Enterococcus faeca
c 22	17	2.1	872	1 X14223	H. pylori GHPO 122
c 23	17	2.1	481	1 V99731	Human adult testis
c 24	17	2.1	110000	1 X20248_08	Continuation (9 of
25	16	2.0	3640	1 N82026	Cloned p2.1 insert
26	16	2.0	1729	1 N81256	EcoRI-BamHI fragme
27	16	2.0	3573	1 N81065	Sequence of human
28	16	2.0	3373	1 N80892	Sequence encoding
29	16	2.0	3440	1 Q13967	Bce-4 gene with re
30	16	2.0	4175	1 Q14099	C-ski protein gene
31	16	2.0	4055	1 Q14799	FB29 chicken c-ski
32	16	2.0	1729	1 N40059	Sequence encoding
33	16	2.0	30	1 Q26869	Primer 1851-M; mut
c 34	16	2.0	1470	1 Q54828	IL-2 receptor gamm
35	16	2.0	26	1 Q72741	Interleukin-2R gam
c 36	16	2.0	2542	1 Q87430	Human cystathionin
37	16	2.0	2870	1 Q91177	Lipase and accesso
38	16	2.0	966	1 Q93034	Wild tomato endoch
39	16	2.0	935	1 Q93035	Wild tomato endoch

c 40 16 2.0 10897 1 T09187 Mutu putative onco  
41 16 2.0 5233 1 T58225 DNA encoding inosi  
c 42 16 2.0 80073 1 T58840\_5 Continuation (6 of  
43 16 2.0 3680 1 T59684 Gene locus for 2.4  
44 16 2.0 5076 1 T59685 Genes for 2.4-diac  
45 16 2.0 6387 1 T59686 Genes for (modulat

ALIGNMENTS

RESULT 1  
X05735  
ID X05735 standard; mRNA; 810 BP.  
AC X05735;  
DT 27-APR-1999 (first entry)  
DE Human beta-amyloid peptide-binding protein (BBP) encoding mRNA.  
KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;  
KW human; Alzheimer's disease; ss.  
OS Homo sapiens.  
FH key Location/Qualifiers  
FT CDS 1..810  
FT /tag= a  
FT /product= "BBP"  
FN WO9846636-A2.  
PD 22-OCT-1998.  
PR 14-APR-1998; U07462.  
PR 16-APR-1997; US-064583.  
PA (AMHP ) AMERICAN HOME PROD CORP.  
PI Bard JA, Jacobsen JS, Kajkowski EM, Ozenberger BA,  
PI Walker SC;  
DR WPI: 99-080736/07.  
DR P-PSDB; W94291.  
PT Polynucleotide encoding beta-amyloid peptide binding protein - used  
PT to identify inhibitors of beta-amyloid peptide for treating  
PT Alzheimer's disease  
PS Cialm 1; Pages 43-44; 59pp; English.  
CC This represents a nucleotide sequence encoding a beta-amyloid peptide  
CC binding protein (BBP). The polynucleotide comprising the entire BBP  
CC nucleotide sequence of clone BBP1-fl is deposited under the accession  
CC number ATCC 98617. The polynucleotide comprising a fragment of BBP  
CC (nucleotides 202-807 of the full length BBP) of clone PEK196 is deposited  
CC as ATCC 98399. Host cells transformed with a vector comprising the BBP  
CC nucleic acid are used for the recombinant production of the protein. The  
CC protein can be used in a method for diagnosing a disease characterised by  
CC aberrant expression of human beta-amyloid protein (BAP). The protein can  
CC also be used in a method for screening for compounds which regulate  
CC expression of a BAP binding protein. The proteins, antibodies and  
CC identified compounds can be used in the treatment or prevention of  
CC Alzheimer's disease.  
SQ Sequence 810 BP; 204 A; 183 C; 202 G; 221 T;

Query Match 100.0%; Score 810; DB 1; Length 810;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 810; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATATTTAAAGGGTCCTCCCAATGTGATTCACGGGCTCAGCGGCAAGAACACG 60  
Db 1 ATGCATATTTAAAGGGTCCTCCCAATGTGATTCACGGGCTCAGCGGCAAGAACACG 60  
QY 61 CGAAGACACGGAAGTGGCCTCTATCTCTATGGAGGTCTCCCTTTAAGAACTCGCCCTGTG 120  
Db 61 CGAAGACACGGAAGTGGCCTCTATCTCTATGGAGGTCTCCCTTTAAGAACTCGCCCTGTG 120  
QY 121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGAAGTGGCGAAGTGTCTGTC 180  
Db 121 CCCTTCTCCCTCCCGCTCCTGGGCGGAGGCGGAAGCGAAGTGGCGAAGTGTCTGTC 180  
QY 181 TCCAAGATGGCGGCGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 240  
Db 181 TCCAAGATGGCGGCGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 240







Qy 166 EVGFKPKISCRNVNGYS---YKVAVALS 190  
 II :: : : I : III : II  
 Db 337 GVQKFYLTGDKNFSRLRYKSVTLS 364

RESULT 13  
 ID P87656 PRELIMINARY; PRT; 644 AA.  
 AC P87656;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE FIBER.  
 OS Eggdrop syndrome-1976 virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae;  
 OC unclassified Adenoviridae.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=127;  
 RA AKOPIAN T.A., DORONIN K.K., ZAKHARCHUK A.N., LOBANOV V.A.,  
 RA NARODITSKY B.S.;  
 RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; 28065; CAB06667.1; -  
 DR PFAM; PF00608; adeno\_fiber2; 4.  
 SQ SEQUENCE 644 AA; 67647 MW; 5A7D86D3 CRC32;

Query Match 5.7%; Score 81.5; DB 12; Length 644;  
 Best Local Similarity 22.08; Pred. No. 20;  
 Matches 3; Conservative 19; Mismatches 53; Indels 59; Gaps

Qy 72 SAPEAVTARLVGLVMSV-----TTGPWGA VATSAGGEESLKCEDLKVGQYI 111  
 II :: : : I : III : II  
 Db 455 SCPQKRAKVAGYVLTSGVGLVHGTHQIKATAGWFT-----GNSVQ-ESIRGLVL 50  
 II :: : : I : III : II  
 Qy 120 CKDKPKINDATQEPVNCNTYTAHVSFPAPNTCKDS-----SGNETHFTGNEVGFFRPI 17  
 I :: : : I : III : II  
 Db 508 C-----PFSARDPTANLSGWPAVVMVSGDSNTPLFYAANAISYTNRVNL----- 55  
 I :: : : I : III : II  
 Qy 174 SCRNVNGYSYKVAVALSLFLGMLGADRYLGYPALGLLKFTCVGFCGI 221  
 I :: : : I : III : II  
 Db 553 ---AVTGNFYKEETEL-----PGYTRHSFCPTGTGM 581  
 I :: : : I : III : II

RESULT 14  
 O11424 PRELIMINARY; PRT; 644 AA.  
 ID O11424  
 AC O11424;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE FIBER PROTEIN.  
 OS avian adenovirus EDS.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=127;  
 RA MEDLINE; 98042451.  
 RX HESS M., BLOECKER H., BRANDT P.;  
 RA "The complete nucleotide sequence of the egg drop syndrome virus: an  
 RT intermediate between mastadenoviruses and aviadenoviruses.";  
 RL Virology 238:145-156(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=127;  
 RA HESS M.;  
 RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y05958; CAA70814.1; -  
 DR PFAM; PF00608; adeno\_fiber2; 4.  
 SQ SEQUENCE 644 AA; 67677 MW; E150260D CRC32;

Query Match 5.7%; Score 81.5; DB 12; Length 644;

RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,  
RA KINASHI H., HOPWOOD D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb streptomycetes coellicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL03443; CAA22367.1; -.  
SQ SEQUENCE 446 AA; 47069 MW; 712D79A5 CRC32;

Query Match 5.8%; Score 84; DB 2; Length 446;  
Best Local Similarity 19.7%; Pred. No. 7.7;  
Matches 52; Conservative 34; Mismatches 110; Indels 68; Gaps 9;  
QY 39 LLPFLPLGGGGGSGEKVSYSKMAAAPSPGSAPEAVTARLVGLWTF--VSVTTPGPG 96  
DB 106 LIVLGVGFGVGGGLGIGYISPTLIKWFDRPG-----MATGIAMGFGGALIASPWS 160  
QY 97 AVATASAGGES-----LKCEDLVGQYICKDPK-----IN 126  
DB 161 AQMLASFGTDSGIALAFVLHGLTYAVFMLGLVLLVRPRERADGRPALEGVQVSAR 220  
QY 127 DATQEP-----VNCNTNTAHVSCF--PAPNITCKDSSGNETHTGNEVGFKPISCRN 177  
DB 221 SAVRTPQFWLLIVLCMNVNTAGIGILEKAPMITDFSTSPSVTAAAGFVALLSAAN 280  
QY 178 VNGYSYKVAVALSLFLGW-----IGADRFYLCYGPALGLLCTGTCGIGSLIDFILIS 231  
DB 281 MAGR-----FGWSASDLGRKNIYRYLVGVGALMTLIALFGDSKPLFVLCA 329  
QY 232 MQIV-----GPSDGSYYIIDYGT 250  
DB 330 LVVVSFYGGGFATAPAYLKDLFT 353

RESULT 9  
P93125  
ID P93125 PRELIMINARY; PRT; 1274 AA.  
AC P93125;  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE TTF-1.  
GN TTF1.  
OS Dunaliella salina.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
CC Dunaliellaceae; Dunaliella.  
RN [1]  
P SEQUENCE FROM N.A.  
RA FISHER M., GOKHMAN I., PICK U., ZAMIR A.;  
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U77059; AAB36531.1; -.  
DR HSP; P56410; LAOV.  
DR PFAM; PF00405; transferrin; 6.  
SQ SEQUENCE 1274 AA; 136667 MW; 1CB5A0D3 CRC32;

Query Match 5.8%; Score 83; DB 10; Length 1274;  
Best Local Similarity 22.9%; Pred. No. 33;  
Matches 30; Conservative 17; Mismatches 50; Indels 34; Gaps 6;  
QY 57 KYSVSKMAAAPSPGSAPEAV-TARLVGLVWFSVTTPGKAVATSAGGESLKCEDLV 115  
DB 579 QDAETIEFWEDNVCAPGSTENGPLIG-----GGKYGEVGENGG----- 620  
QY 116 GYICKDPKINDATQEPVNCNTNTAHVSCFPAPNITCKDSSGNETHTGNEVGFKPIS 175  
DB 620 ---LCKRCRTDCTSEDPY--AGYDGAHVCI-----DDDDGNQ--FTGDIADFVHSTL 665  
QY 176 RNVNGYSYKVA 186  
DB 666 RYNGPNLNTA 676

RESULT 10  
Q9XBP8  
ID Q9XBP8 PRELIMINARY; PRT; 789 AA.  
AC Q9XBP8;  
DT 01-NOV-1999 (TREMELrel. 12, Created)  
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE SERINE/THREONINE KINASE PKN11.  
GN PKN11.  
OS Myxococcus xanthus.  
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DZF1.  
RA INOUE S., JAIN R., UEKI T., NARIYA H., XU C., HSU M.,  
RA MUNOZ-DORADO J., FAREZ-VIDAL E., INOUE M.;  
RT "Sequence Analysis of 13 Eukaryotic-like Protein Ser/Thr Kinases of  
Myxococcus xanthus, a Developmental Bacterium and Significance of  
RT Their Coexistence with Protein His Kinases.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF159690; AAD42854.1; -.  
KW Kinase.  
SQ SEQUENCE 789 AA; 81613 MW; FB7C3265 CRC32;

Query Match 5.7%; Score 82.5; DB 2; Length 789;  
Best Local Similarity 26.6%; Pred. No. 21;  
Matches 49; Conservative 16; Mismatches 76; Indels 43; Gaps 8;  
QY 5 KGSP-NVIPRAHGOKNTRDGTGLYMRGPFKNLALLP-----FSLPLLG 48  
DB 500 EGTPTNTEPEAHPCMRAPGSEVPAAGAHASGAEPAASARASAGRAHAFSGVLG 559  
QY 49 GGGSGSGEKY-----SVSKMAAAPSPGSAPEAVTARLVGLVWFSVTTPG---PWGA 97  
DB 560 TSARGSGAEVPAGAHASGAELEAGARASGPVASVAVPAARRQWAMGLAVGLAVVGG 619  
QY 98 VATSAGEESLKCEDLVGQYICKDPKI-NDATQEPVNCNTNTAHVSCFPAPNITCKDSS 156  
DB 620 VAVMRGTEEA-----APPAVPAEPMPVAAPAVDPEVP-----TAQ-----PAPSAVSENAE 664  
QY 157 GNET 160  
DB 665 GGET 668

RESULT 11  
P93368  
ID P93368 PRELIMINARY; PRT; 409 AA.  
AC P93368;  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE DIPHENOL OXIDASE (EC 1.10.3.2) (LACCASE) (URISHTOL OXIDASE)  
DE (FRAGMENT).  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
OC Nicotiana.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 97080572.  
RA KIEFER-MEYER M., GOMORD V., O'CONNELL A., HALPIN C., FAYE L.;  
RT "Cloning and sequence analysis of laccase-encoding cDNA clones from  
tobacco.";  
RL Gene 178:205-207(1996).  
CC -I- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE + 2  
H(2)O.  
CC -I- COFACTOR: COPPER.  
DR EMBL; U45243; AAC49538.1; -.  
DR MENDEL; 9402; Nicta;1457;9402.





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2000, 04:33:16 : Search time 14.98 Seconds  
(without alignments)  
1245.052 Million cell updates/sec

Title: US-09-060-609-2  
Perfect score: 1439  
Sequence: 1 MHILKSPNVIPRAHGQKNT.....TRLRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62

Searched: 225878 seqs, 69334122 residues

Database: SPTREMBL\_12.\*

Word size: 0

Number of hits that pass the threshold : 225878

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	95.5	6.6	913	4	O75061	O75061 homo sapien
2	92	6.4	456	12	O66668	O66668 equine herp
3	89.5	6.2	690	10	O22512	O22512 glycine max
4	89	6.2	330	12	O98339	O98339 murine hepa
5	88	6.1	1324	12	O39227	O39227 murine hepa
6	87.5	6.0	1207	5	Q21535	Q21535 caenorhabdi
7	86	6.0	132	11	O92337	O92337 mus musculus
8	84	5.8	446	2	Q92BV5	Q92BV5 streptomyc
9	83	5.8	1274	10	P93125	P93125 dunalieila
10	82.5	5.7	789	2	Q9XB98	Q9XB98 myxococcus
11	82	5.7	409	10	P93368	P93368 nicotiana t
12	82	5.7	465	11	O88354	O88354 spermophili
13	81.5	5.7	644	12	P87656	P87656 eggdrops syn
14	81.5	5.7	644	12	O11424	O11424 avian adeno
15	81.5	5.7	3432	12	O90297	O90297 japanese en
16	80.5	5.6	455	4	Q99808	Q99808 homo sapien
17	80.5	5.6	377	5	Q9XY68	Q9XY68 balanus bal
18	80.5	5.6	3432	12	O90417	O90417 japanese en
19	80.5	5.6	1237	13	O91976	O91976 gallus gall
20	80	5.6	638	5	O77408	O77408 lymnaea sta
21	80	5.6	171	10	O81433	O81433 arabidopsis
22	79.5	5.5	489	2	O33261	O33261 mycobacteri
23	79.5	5.5	171	9	Q38215	Q38215 insertion s
24	79.5	5.5	530	12	Q83876	Q83876 norway vir
25	79.5	5.5	3432	12	O82873	O82873 japanese en

26	79.5	5.5	3432	12	O82872	O82872 japanese en
27	79.5	5.5	3432	12	O9WCX7	O9WCX7 japanese en
28	79.5	5.5	1064	12	Q9WAU6	Q9WAU6 broad bean
29	79	5.5	1278	4	O60434	O60434 homo sapien
30	79	5.5	797	4	O95288	O95288 homo sapien
31	79	5.5	3507	5	Q23587	Q23587 caenorhabdi
32	78.5	5.5	482	2	Q92MT7	Q92MT7 helicobacte
33	78.5	5.5	7962	4	Q10465	Q10465 homo sapien
34	78.5	5.5	5376	11	O88799	O88799 mus musculu
35	78	5.4	419	2	P72207	P72207 pseudomonas
36	78	5.4	523	10	Q39438	Q39438 beta vulgar
37	78	5.4	3019	12	O92529	O92529 hepatitis c
38	78	5.4	414	12	O81267	O81267 hepatitis c
39	78	5.4	446	13	Q9YIB2	Q9YIB2 anguilla ja
40	77.5	5.4	374	1	O26843	O26843 methanobact
41	77.5	5.4	869	5	Q9Y028	Q9Y028 drosophila
42	77	5.4	797	4	Q9Y345	Q9Y345 homo sapien
43	77	5.4	1239	5	O61541	O61541 drosophila
44	77	5.4	1302	5	O61542	O61542 drosophila
45	77	5.4	1049	10	Q9ZNR9	Q9ZNR9 volvox cart

#### ALIGNMENTS

RESULT 1  
O75061 PRELIMINARY; PRT; 913 AA.  
AC O75061;  
DT 01-NOV-1998 (TREMREL. 08, Created)  
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)  
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)  
DE KIAA0473 PROTEIN.  
GN KIAA0473.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE; 98116662.  
RA SEKI N., OHIRA M., NAGASE T., ISHIKAWA K., MIYAJIMA N., NAKAJIMA D.,  
RA NOMURA N., OHARA O.;  
RT "Characterization of cDNA clones in size-fractionated cDNA libraries  
RT from human brain."  
RL DNA Res. 4:345-349(1997).  
DR EMBL; AB007942; BRA32318.1; --  
SQ SEQUENCE 913 AA; 99996 MW; 025A0952 CRC32;

Query Match 6.6% Score 95.5; DB 4; Length 913;  
Best Local Similarity 24.9%; Pred. No. 1.7;  
Matches 48; Conservative 23; Mismatches 87; Indels 35; Gaps 8;  
QY 29 PMRGPKNLALPFLSLLGGGGSGGKSVSKMAAAW-PSGSPAPEAVTAR----- 81  
Db 513 PRAAPTNSSELLS---DLFGGGGAAGTQAGQSGVEDVHFPGSPASTQSTPRRSATSTSA 569  
QY 81 ----LVGVLFVSVTTGPGCAVATSGAGGEEELKCEDLKVGQYICKDKPKINDATQEPVNCT 136  
Db 570 SPTLRVG---EGATDPFGAPSKPSG-----QDL-LGSFLNTSSASSDPFLQPTRSP 617  
QY 137 NYTAHVSCEFAPNITCKDSNGNETHFTGNEVGFKPKISCRNVNGYSYKAV-----ALS 190  
Db 618 SPTVHASSTPAVNIQPDVSGGWDHAKPGFGMGSKSAATSPTGSSHGTTTHOSKPQTLD 677  
QY 191 LF--LGLWLGADRF 201  
Db 678 PFADLTGLSSSF 690  
RESULT 2  
Q66668

Search completed: February 11, 2000, 03:33:27  
Job time: 3116 sec



```
208 GGTCCGCTCTCCGAGGCGCGTACGCCAGACTCGTGTGTCCTGTCGTCTCA 267
Db 1 GGTCCGCTCTCCGAGGCGCGTACGCCAGACTCGTGTGTCCTGTCGTCTCA 60
QY 268 GTCACTACAGGACCTGGGGGGGTGTTCACCTCCCGGGGGCGAGGAGTGCCTTAAG 327
Db 61 GTCACTACAGGACCTGGGGGGGTGTTCACCTCCCGGGGGCGAGGAGTGCCTTAAG 120
QY 328 TCGAGGACCTCAAGTGGGACATATATTGTAAGATCCAAATAAATGAGCTACG 387
Db 121 TCGAGGACCTCAAGTGGGACATATATTGTAAGATCCAAATAAATGAGCTACG 180
QY 388 CAAGAACCAGTAACTGACAACTACACAGCTCATGTTTCCTGTTTCCAGACCCCAAC 447
Db 181 CAAGAACCAGTAACTGACAACTACACAGCTCATGTTTCCTGTTTCCAGACCCCAAC 240
QY 448 ATAACTGTAAGGATCCAGTGGCAATGAACACATTTTACTGGGAACGAAGTGGTTTT 507
Db 241 ATAACTGTAAGGATCCAGTGGCAATGAACACATTTTACTGGGAACGAAGTGGTTTT 300
QY 508 TTCAAGCCCATATCTGCCAATGTAATGGCTATTCCTTACAAAGTGGCAGTCGATTG 567
Db 301 TTCAAGCCCATATCTGCCAATGTAATGGCTATTCCTTACAAAGTGGCAGTCGATTG 360
QY 568 TCTCTTTTCTTGATGGTGGGAGCAGATCGATTTTACCTTGGATACCTGCTTTGGGT 627
Db 361 TCTCTTTTCTTGATGGTGGGAGCAGATCGATTTTACCTTGGATACCTGCTTTGGGT 420
QY 628 TTGTTAAAGTTTGGCACTG 646
Db 421 TTGTTAAAGTTTGGCACTG 439

RESULT 13
AI038331
LOCUS
DEFINITION
ox84a06.xl Soares_senescent_fibroblasts_NBHSF Homo sapiens cDNA
clone IMAGE:1662994 3', mRNA sequence.
ACCESSION
AI038331
VERSION
AI038331.1 GI:3277525
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1099 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 432.
Location/Qualifiers
1..448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1662994"
/clone_lib="Soares_senescent_fibroblasts_NBHSF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="vector: pT73D (Pharmacia) with a modified
polylinker V_Type: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCATCTGAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3']
```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo.\*

BASE COUNT  
118 a 92 c 105 g 133 t

Query Match 54.1%; Score 438.2; DB 41; Length 448;  
Best Local Similarity 99.3%; Pred. No. 5e-108;  
Matches 440; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 258 GTTCGTCTCAGTCACTACAGACCTGGGGGGCTGTTCGCCACCTCCCGGGGGCGAGA 317
Db 1 GATCGTCTCAGTCACTACAGACCTGGGGGGCTGTTCGCCACCTCCCGGGGGCGAGA 60
QY 318 GTCGCTTAAAGTGGGAGGACCTCAAGTGGGACATATATTTGTAAGATCCAAAAATAAA 377
Db 61 GTCGCTTAAAGTGGGAGGACCTCAAGTGGGACATATATTTGTAAGATCCAAAAATAAA 120
QY 378 TGAGCTACGCAAGAACAGTTAACTGTACAACTACACAGCTCATGTTCTGTTTCC 437
Db 121 TGAGCTACGCAAGAACAGTTAACTGTACAACTACACAGCTCATGTTCTGTTTCC 180
QY 438 AGCACCCCAACATACTTGTAGGATCCAGTGGCAATGAACACACATTTTACTGGGAACGA 497
Db 181 AGCACCCCAACATACTTGTAGGATCCAGTGGCAATGAACACACATTTTACTGGGAACGA 240
QY 498 AGTTGGTTTTTCAAGCCCATATCTTGCCGAAATGTAATGGCTATTCCTCAAAAGTGC 557
Db 241 AGTTGGTTTTTCAAGCCCATATCTTGCCGAAATGTAATGGCTATTCCTCAAAAGTGC 300
QY 558 AGTCGATTTCTCTTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTTGGATACCC 617
Db 301 AGTCGATTTCTCTTTTCTTGGATGGTGGGAGCAGATCGATTTTACCTTGGATACCC 360
QY 618 TGCTTTGGTTTGTAAAGTTTGCACGTAGAGTTTGTGGAATGGGAGCTTAATGA 677
Db 361 TGCTTTGGTTTGTAAAGTTTGCACGTAGATGTTTGTGGAATGGGAGCTTAATGA 420
QY 678 TTTCATTTCTATTCAATGCAGA 700
Db 421 TTTCATTTCTATTCAATGCAGA 443
```

RESULT 14

AI168073  
LOCUS  
DEFINITION  
ox89c02.xl Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA  
clone IMAGE:1682498 3', mRNA sequence.  
ACCESSION  
AI168073  
VERSION  
AI168073.1 GI:3701243  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 471)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2282128.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1..471

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1--(bases 1 to 475)  
 AUTHORS Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,  
 Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,  
 White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,  
 Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,  
 Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geoghagen N.S.,  
 Glodek A., Gnehm C.D., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,  
 Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,  
 Moreno-Palancas R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,  
 Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R.,  
 Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li, Y.,  
 Bebnarik D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,  
 Dimke D., Feng D.-F., Ferrie A., Fischer C., Hastings G.A.,  
 He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,  
 Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,  
 Raymond L., Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M.,  
 Dillion P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,  
 Fraser C.M. and Venter J.C.  
 TITLE Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 12140200  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288147.  
 Other\_ESTs: THCI95202  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (http://www.tigr.org/tgdb/hgi/hgi.html)  
 Seq primer: M13 Reverse.

FEATURES  
 Source  
 1. 475  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):108677"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Colon carcinoma (HCC) cell line"  
 /tissue\_type="colon"  
 /cell\_type="KM12C"  
 /cell\_line="KM12C(HCC)-parental human colon  
 carcinoma/Dukes B2"  
 /note="Organ: colon; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI"  
 BASE COUNT 108 a 107 c 119 g 135 t 6 others  
 ORIGIN  
 Query Match 54.9%; Score 444.4; DB 31; Length 475;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-109;  
 Matches 467; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 189 GCGGGCGCCCTGGCGTGTGGTCGCTGCTCCGAGGCGCTGAGCGCCAGACTCGTTGG 248  
 Db 1 GCGGGCGCCCTGGCGTGTGGTCGCTGCTCCGAGGCGCTGAGCGCCAGACTCGTTGG 59  
 QY 249 TGTCCTGTGGTGTCTCAGTCACAGGACCTCGGGGCTGTGTCACCTCCGCGGG 308  
 Db 60 TGTCCTGTGGTGTCTCAGTCACAGGACCTCGGGGCTGTGTCACCTCCGCGGG 119  
 QY 309 GCGGAGGAGTCGCTTAAGTCGCGAGGACCTCAAGTCGGGACCAATATTTGTAAGATCC 368  
 Db 120 GCGGAGGAGTCGCTTAAGTCGCGAGGACCTCAAGTCGGGACCAATATTTGTAAGATCC 179  
 QY 369 AAAATAAATACGCTACGACGACCAAGTAACTGTACAACTACACAGCTCATGTTTC 428  
 Db 180 AAAATAAATACGCTACGACGACCAAGTAACTGTACAACTACACAGCTCATGTTTC 239

QY 429 CTGTTTCCAGCACCAACATACTTCTAAGGATTCAGTGGCAATGAACACATTTTAC 488  
 Db 240 CTGTTTCCAGCACCAACATACTTCTAAGGATTCAGTGGCAATGAACACATTTTAC 299  
 QY 489 TGGGACGAGGTTGGTTTTTTCAGCCCATATCTTCGCGAAATCTAAATGCTATTCCTA 548  
 Db 300 TGGGACGAGGTTGGTTTTTTCAGCCCATATCTTCGCGAAATCTAAATGCTATTCCTA 359  
 QY 549 CAAAGTGGCAGTCGCGATGCTCTTTTCTTGGATGTTGGGACGAGATCGATTTTACCT 608  
 Db 360 CAAAGTGGCAGTCGCGATGCTCTTTTCTTGGATGTTGGGACGAGATCGATTTTACCT 419  
 QY 609 TGGATACCTGCTTTGGTTGTTAAAG-TTTTGCAGTGTAGGGTTTGTGGAATT 663  
 Db 420 TGGATACCTGCTTTGGTTGTTAAAGTTTTCAGTGTAGGGTTTGTGGAATT 475

RESULT 12  
 LOCUS AI299154 439 bp mRNA EST 29-JAN-1999  
 DEFINITION qm96f11.xl NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1896621 3',  
 mRNA sequence.  
 ACCESSION AI299154  
 VERSION AI299154.1 GI:39588808  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 439)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 966 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 419.  
 Location/Qualifiers  
 1. 439  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1896621"  
 /clone\_lib="NCI\_CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 neuroendocrine lung carcinoid, and was then primed with a  
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
 to Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT73 vector. Library is normalized. Library was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 105 a 100 c 110 g 124 t  
 ORIGIN

Query Match 54.2%; Score 439; DB 44; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 3e-108;  
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 1 GCGGCGGCTGGCGCTGGTCCGCTCTCGCGAGGCTTGACGCCAGACTCGTTGGT 60
Qy 250 GTCCTGTGCTGCTCAGTCACTACAGACCTCGGGGGCTGTGGCCACCTCCGCCGGG 309
Db 61 GTCCTGTGCTGCTCAGTCACTACAGACCTCGGGGGCTGTGGCCACCTCCGCCGGG 120
Qy 310 GCGGAGGAGTGGCTTAAGTGGGAGGACCTCAAGTGGGACAATATATTGTAAGATCCA 369
Db 121 GCGGAGGAGTGGCTTAAGTGGGAGGACCTCAAGTGGGACAATATATTGTAAGATCCA 180
Qy 370 AAATAAATGAGCTGACGACAGACCTTAAGTGGGACAATATATTGTAAGATCCA 429
Db 181 AAATAAATGAGCTGACGACAGACCTTAAGTGGGACAATATATTGTAAGATCCA 240
Qy 430 TGTTCCTCAGCACCACCAATCACTGTAAGGATTCAGTGGCAATGAACACATTTTACT 489
Db 241 TGTTCCTCAGCACCACCAATCACTGTAAGGATTCAGTGGCAATGAACACATTTTACT 300
Qy 490 GCGAACGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 549
Db 301 GCGAACGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 360
Qy 550 AAGTGGCAGTCGCAATGCTCTTTTCTTGAGAGTGGGAGCAGATCGATTTTACCTT 609
Db 361 AAGTGGCAGTCGCAATGCTCTTTTCTTGAGAGTGGGAGCAGATCGATTTTACCTT 420
Qy 610 GGATACCTGCTTGGGTTTGTAAAGTTTTCAGTCTAGGTTTGTGGAATGGGAGC 669
Db 421 GGATACCTGCTTGGGTTTGTAAAGTTTTCAGTCTAGGTTTGTGGAATGGGAGC 480
Qy 670 CTAATTGATTTCACTTCTTAT 689
Db 481 CTAATTGATTTCACTTCTTAT 500
```

```
RESULT 10
N47594
LOCUS N47594 507 bp mRNA EST 14-FEB-1996
DEFINITION Y59a07.sl Soares_multiple_sclerosis_2NbHMP Homo sapiens cDNA
clone IMAGE:277812 3', mRNA sequence.
ACCESSION N47594
VERSION N47594.1 GI:1188760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 507)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On May 9, 1995 this sequence version replaced gi:802394.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: ml3 -40 forward
High quality sequence stop: 459.
Location/Qualifiers
1. .507
/organism="Homo sapiens"
/db_xref="GDB:3896188"
/db_xref="taxon:9606"
FEATURES
source
```

```
/clone="IMAGE:277812"
/clone_lib="Soares_multiple_sclerosis_2NbHMP"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: p7T3D (Pharmacia) with a modified
polylinker V-type: phagemid; Site.1: Not I; Site.2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCAGAGTGGGAGCGGCCCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 117 a 109 c 124 g 148 t 9 others
ORIGIN
Query Match 56.0%; Score 453.2; DB 25; Length 507;
Best Local Similarity 95.3%; Pred. No. 4.6e-112;
Matches 484; Conservative 0; Mismatches 21; Indels 3; Gaps 2;
Qy 190 GCGGCGGCTGGCGCTGGTCCGCTCTCGCGAGGCGCTGACGCCAGACTCGTTGGT 249
Db 1 GCGNCGGCTGCGTGGTCCGCTCTCGCGAGGCGCTGACGCCAGACTCGTTGGT 58
Qy 250 GTCCTGTGCTGCTCAGTCACTACAGACCTCGGGGGCTGTGGCCACCTCCGCCGGG 309
Db 59 GTCCTGTGCTGCTCAGTCACTACAGACCTCGGGGGCTGTGGCCACCTCCGCCGGG 118
Qy 310 GCGGAGGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 369
Db 119 GNNAGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 178
Qy 370 AAATAAATGAGCTGACGACAGACCTTAAGTGGGACAATATATTGTAAGATCCA 429
Db 179 AAATAAATGAGCTGACGACAGACCTTAAGTGGGACAATATATTGTAAGATCCA 238
Qy 430 TGTTCCTCAGCACCACCAATCACTGTAAGGATTCAGTGGCAATGAACACATTTTACT 489
Db 239 TGTTCCTCAGCACCACCAATCACTGTAAGGATTCAGTGGCAATGAACACATTTTACT 298
Qy 490 GCGAACGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 549
Db 299 GCGAACGAGTGGTGTTCCTCAAGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 358
Qy 550 AAGTGGCAGTCGCAATGCTCTTTTCTTGAGAGTGGGAGCAGATCGATTTTACCTT 609
Db 359 AAGTGGCAGTCGCAATGCTCTTTTCTTGAGAGTGGGAGCAGATCGATTTTACCTT 418
Qy 610 GGATACCTGCTTGGGTTTGTAAAGTTTTCAGTCTAGGTTTGTGGAATGGGAGC 668
Db 419 GGATACCTGCTTGGGTTTGTAAAGTTTTCAGTCTAGGTTTGTGGAATGGGAGC 478
Qy 669 CCTAATTGATTTCACTTCTTATTTCAATG 696
Db 479 CCTAATTGATTTCACTTCTTATTTCAATG 506
$ RESULT 11
AA306979
LOCUS AA306979 475 bp mRNA EST 18-APR-1997
DEFINITION EST178050 Colon carcinoma (HCC) cell line Homo sapiens cDNA 5' end,
mRNA sequence.
ACCESSION AA306979
VERSION AA306979.1 GI:1959469
KEYWORDS EST.
SOURCE human.
```

AI349520 498 bp mRNA EST 03-FEB-1999  
LOCUS qp72905.xl Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
DEFINITION IMAGE:1928600 3', mRNA sequence.  
ACCESSION AI349520  
VERSION AI349520.1 GI:4086726  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 498)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
Insert Length: 1009 Std Error: 0.00  
Seq primer: 40UP from Gibco.  
FEATURES  
Location/Qualifiers  
1..498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1928600"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NbHL19W."  
BASE COUNT 113 a 112 c 128 g 144 t 1 others  
ORIGIN  
Query Match 61.4%; Score 497; DB 45; Length 498;  
Best Local Similarity 99.8%; Pred. No. 6.7e-124;  
Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 190 GGGCGCGCTGCGCTGCTCGTCTCGGAGCGCGTGACGCCAGACTCGTTGGT 249  
Db 1 GCGGCGCGCTGCGCTGCTCGTCTCGGAGCGCGTGACGCCAGACTCGTTGGT 60  
QY 250 GTCCTGTGTTGCTCTAGTCACTACAGGACCTCGGGGGCTGTTCACCTCCGCGGG 309  
Db 61 GTCCTGTGTTGCTCTAGTCACTACAGGACCTCGGGGGCTGTTCACCTCCGCGGG 120  
QY 310 GCGAGGAGTGCCTTAAGTCGAGGACCTCAAAAGTGGGCAATATATTTGTAAGATCCA 369  
Db 121 GCGAGGAGTGCCTTAAGTCGAGGACCTCAAAAGTGGGCAATATATTTGTAAGATCCA 180  
QY 370 AAAATAATGACGCTACCGAAGACAGTTACTGTACAACTACACGCTCATGTTTC 429  
Db 181 AAAATAATGACGCTACCGAAGACAGTTACTGTACAACTACACGCTCATGTTTC 240  
QY 430 TCTTTTCCAGCACCAACAATCTGTAGGATTCAGTGGCAATGAACACATTTTACT 489  
Db 241 TCTTTTCCAGCACCAACAATCTGTAGGATTCAGTGGCAATGAACACATTTTACT 300  
QY 490 GGAACGAGTGGTGGTTTTTCAAGCCCATATCTTGCAGAAATGTAATGGCTATTCCTAC 549  
|||||

Db 301 GGGAAACAGTTGCTTTTTCACGCCCATATCTTCCGAAATGTAATGGCTATTCCTAC 360  
QY 550 AAGTGGCAGTCGCAATGCTCTCTTTTCTTGATGGTTGGGAGCAGATCGATTACCTT 609  
|||||  
Db 361 AAGTGGCAGTCGCAATGCTCTCTTTTCTTGATGGTTGGGAGCAGATCGATTACCTT 420  
QY 610 GGATACCTGCTTTGGCTTGTAAAGTTTTCACACTGTAGGTTTGTGGAATTGGGAGC 669  
|||||  
Db 421 GGATACCTGCTTTGGCTTGTAAAGTTTTCACACTGTAGGTTTGTGGAATTGGGAGC 480  
QY 670 CTAATTCATTTTCATTCCT 687  
|||||  
Db 481 CTAATTCATTTTCATTCCT 498  
RESULT 9  
AI680969  
LOCUS 500 bp mRNA EST 26-MAY-1999  
DEFINITION tx35807.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2271613 3',  
mRNA sequence.  
ACCESSION AI680969  
VERSION AI680969.1 GI:4891151  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 500)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135321.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 500.  
FEATURES  
Location/Qualifiers  
1..500  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2271613"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI\_CGAP\_Lu5 was prepared, and ss circles were  
used in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 114 a 112 c 125 g 145 t 4 others  
ORIGIN  
Query Match 60.8%; Score 492.8; DB 50; Length 500;  
Best Local Similarity 98.8%; Pred. No. 9.2e-123;  
Matches 494; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 190 GCGGCGCGCTGCGCTGCTCGTCTCGGAGCGCGTGACGCCAGACTCGTTGGT 249







LOCUS AI680904 574 bp mRNA EST 26-MAY-1999  
DEFINITION tx42f05.x1 NCI\_CGAP.Lu24 Homo sapiens cDNA clone IMAGE:2272257 3',  
similar to WP:CO2F5.3 CE00039 GTP-BINDING PROTEIN ;, mRNA sequence.  
ACCESSION AI680904  
VERSION AI680904.1 GI:4891086  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 574)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1134035.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael P.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbbrp/image/image.html](http://www-bio.llnl.gov/dbbrp/image/image.html)

Seq primer: -40UP from Gibco  
High quality sequence stop: 459.  
Location/Qualifiers  
1..574

ORGANISM Homo sapiens  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2272257"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI\_CGAP.Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
PAGE COUNT 148 a 118 c 136 g 172 t

Query Match 70.7%; Score 572.4; DB 50; Length 574;  
Best Local Similarity 99.8%; Pred. No. 3e-144;  
Matches 573; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 211 CGGTCTGCTCCGGAGGCCGTGACGCCAGACTCGTTGTGTCCTGTGTTTCGTTCTCAGTC 270  
Db 1 CGGTCTGCTCCGGAGGCCGTGACTGCCAGCTGCTGTGTCCTGTGTTTCGTTCTCAGTC 60  
QY 271 ACTACAGGACCTCGGGGGCTGTGTCACCTCCGGGGGGGGAGAGTCGCTTAAGTGC 330  
Db 61 ACTACAGGACCTCGGGGGCTGTGTCACCTCCGGGGGGGGAGAGTCGCTTAAGTGC 120  
QY 331 GAGGACCTCAAGTGGGACAATATTTCTAAGATCCAAAATAAATCAACCTACGCCA 390  
Db 121 GAGGACCTCAAGTGGGACAATATTTCTAAGATCCAAAATAAATCAACCTACGCCA 180  
QY 391 GAAACAGTAACTGTACAAACTACAGCTCATGTTCTGTTTCCAGCACCCACACATA 450  
Db 181 GAACAGTAACTGTACAAACTACAGCTCATGTTCTGTTTCCAGCACCCACACATA 240  
QY 451 ACTTGAAGGATTCAGTGGCAATGAACACACATTTTACTGGGAACGAAGTTGTTTTTC 510

Db 241 ACTTGTAAGGATTCAGTGGCAATGAACACACATTTTACTGGGAACGAAGTTGGTTTTTC 300  
QY 511 AAGCCCATATCTTCCCGAATGTAATGGGTATTCCTACAAAGTGGCAGTCGCAATTGTCT 570  
Db 301 AAGCCCATATCTTCCCGAATGTAATGGGTATTCCTACAAAGTGGCAGTCGCAATTGTCT 360  
QY 571 CTTTTTCTTGATGGTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTGGGTTTG 630  
Db 361 CTTTTTCTTGATGGTGGGAGCAGATCGATTTTACCTTGGATACCCCTGCTTGGGTTTG 420  
QY 631 TTAAGTTTTGCACCTAGAGGTTTTGTGGAATGGGAGCCCTAAATTGATTTTCATTTATT 690  
Db 421 TTAAGTTTTGCACCTAGAGGTTTTGTGGAATGGGAGCCCTAAATTGATTTTCATTTATT 480  
QY 691 TCAATGCAGATTTGTGGACCTTCAGATGGAAGTAGTTACATTATATAGATTACTGGAACC 750  
Db 481 TCAATGCAGATTTGTGGACCTTCAGATGGAAGTAGTTACATTATATAGATTACTGGAACC 540  
QY 751 AGACTTACAGACTGAGTATTAATGAACAT 784  
Db 541 AGACTTACAGACTGAGTATTAATGAACAT 574

RESULT 3  
AI479764 556 bp mRNA EST 14-APR-1999  
LOCUS tm69b04.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2163343 3',  
DEFINITION mRNA sequence.  
ACCESSION AI479764  
VERSION AI479764.1 GI:4372932  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 556)  
AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRGAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188819.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbbrp/image/image.html](http://www-bio.llnl.gov/dbbrp/image/image.html)

Insert Length: 614 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 456.  
Location/Qualifiers  
1..556  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2163343"  
/clone\_lib="NCI\_CGAP\_Brn25"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGGAGCGCCGATAGGTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.

FEATURES  
source







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 11, 2000, 02:41:31 ; Search time 506.49 Seconds  
(without alignments)  
6038.200 Million cell updates/sec

Title: US-09-060-609-1  
Perfect score: 810  
Sequence: 1 ATGCATATTTAAAGGTC.....AAACGCATATATCCATAA 810

Scoring table: IDENTITY\_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:\*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*

50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	595.4	73.5	599	62	AI923178 wt67b10.x
2	572.4	70.7	574	50	AI680904 tx42f05.x
3	555	68.5	556	47	AI479764 tm69b04.x
4	531.2	65.6	537	48	AI580361 tm47h02.x
5	527.4	65.1	530	50	AI674462 wc44e01.x
6	513.4	63.4	548	38	AA772225 ai4lc01.s
7	506	62.5	507	50	AI682204 wa71b06.x